



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 163760**

**TO: Patricia Duffy**  
**Location: 3b05 / 3c18**  
**Sunday, August 28, 2005**  
**Art Unit: 1645**  
**Phone: 571-272-0855**  
**Serial Number: 10 / 078531**

**From: Jan Delaval**  
**Location: Biotech-Chem Library**  
**Remsen 1a51**  
**Phone: 571-272-2504**  
**jan.delaval@uspto.gov**

### **Search Notes**

**THIS PAGE BLANK (USPTO)**



From: Duffy, Patricia  
Sent: Thursday, August 25, 2005 8:37 AM  
To: STIC-Biotech/ChemLib  
Subject: sequence search

In re: 10/078,531

Please search SEQ ID NO:2 and oligomers thereof.  
Please include both a sequence and interference search.  
Please print out top 100 hits in each category.

Thank you.

Patricia A. Duffy, Ph.D.  
Art Unit 1645  
Remsen 3B05; Mailbox 3C18  
571-272-0855

\*\*\*\*\*

STAFF USE ONLY

Searcher: [Signature]  
Searcher Phone: 2- 2504  
Date Searcher Picked up: 8/24/05  
Date Completed: 8/24/05  
Searcher Prep/Rev. Time: 20  
Online Time: 15

\*\*\*\*\*

Type of Search

NA#: \_\_\_\_\_ AA#: 2  
Interference: ✓ SPDI: \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: ✓  
Encode/Transl: \_\_\_\_\_  
Structure#: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: ✓  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**



105 197 3.9 2441 2 096124 O96124 plasmodium  
106 197 3.9 4007 2 075H24 Q75H24 neurospora  
107 196.5 3.9 1680 2 09U8G1 Q9U8G1 plasmodium  
108 196.5 3.9 1875 1 MLF1\_YEAST Q04551 saccharomyc  
109 196.5 3.9 2139 2 Q07569 Q07569 entamoeba h  
110 196.5 3.9 2479 2 Q66GS8 Q66GS8 homo sapien

ALIGNMENTS

RESULT 1  
Q9A0C0  
ID Q9A0C0 PRELIMINARY; PRT; 1008 AA.  
AC Q9A0C0  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Hypothetical protein SPY0843.  
GN OrderedLocustNames=SPY0843;  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;  
RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Sezzate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).  
DR EMBL; AE006534; AAK33772.1; -  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR007093; LRR\_Tp.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 1008 AA; 111503 MW; 045BF8CE931AF0CF CRC64;

Query Match 100.0%; Score 5080; DB 2; Length 1008;  
Best Local Similarity 100.0%; Pred. No. 2.4e-215; Indels 0; Gaps 0;  
Matches 1008; Conservative 0; Mismatches 0;

QY 1 MKKHLKLTVALTTLTVSVTHNQEVFSLVKEPIKQTSASSISGADYAESGSKLKINE 60  
DB 1 MKKHLKLTVALTTLTVSVTHNQEVFSLVKEPIKQTSASSISGADYAESGSKLKINE 60  
QY 61 TSGPVDVDTVDLPSDKRTTPEIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120  
DB 61 TSGPVDVDTVDLPSDKRTTPEIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120  
QY 121 SLSLNTKTPSTNWEICDPTTKGNTLVGLSKGVKLSQTDHLVLPSSQAADTQLIQVAS 180  
DB 121 SLSLNTKTPSTNWEICDPTTKGNTLVGLSKGVKLSQTDHLVLPSSQAADTQLIQVAS 180  
QY 181 FAFTPDKKTAAIAYTSRAGENGIEISQLDVKGKEIINEGEVFNLSYLLKKVTIPTGYKHIGQ 240  
DB 181 FAFTPDKKTAAIAYTSRAGENGIEISQLDVKGKEIINEGEVFNLSYLLKKVTIPTGYKHIGQ 240  
QY 241 DAFVFNKNTAFVNLPSLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITKLSIP 300  
DB 241 DAFVFNKNTAFVNLPSLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITKLSIP 300  
QY 301 RQMLAEAFKSNHKTTEFGNSLKVIGEASFQNDLSQMLPDGLEKIESAFTGNP 360  
DB 301 RQMLAEAFKSNHKTTEFGNSLKVIGEASFQNDLSQMLPDGLEKIESAFTGNP 360  
QY 361 GDDHYNNRVVLTKGKNPSGLATENTVYNPKSLWQESPEIDYTKWLEEDFTYQKNSVT 420  
DB 361 GDDHYNNRVVLTKGKNPSGLATENTVYNPKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

QY 421 GFSNKGLOKVRKNKLEIPKOHNGVTTIEIGDNAPRVNDFONKTLRKVDLEEVKLPSTIR 480  
DB 421 GFSNKGLOKVRKNKLEIPKOHNGVTTIEIGDNAPRVNDFONKTLRKVDLEEVKLPSTIR 480  
QY 481 KIGAFAPQSNNLKSFEASDDLEETKEGAFMNNRLETLELKDQKLVITIGDAAPHINHIYAIV 540  
DB 481 KIGAFAPQSNNLKSFEASDDLEETKEGAFMNNRLETLELKDQKLVITIGDAAPHINHIYAIV 540  
QY 541 LPESVQIEIGRSAPFRONGANNLI FMGSKVKTIGEMAFISNRLEHLDLSEQKOLTPIVQAF 600  
DB 541 LPESVQIEIGRSAPFRONGANNLI FMGSKVKTIGEMAFISNRLEHLDLSEQKOLTPIVQAF 600  
QY 601 SDNALKEVLLPASLKTITREAFKKNHLKQLEVASALSASHIAENALDDNDGDGDFONKVVVK 660  
DB 601 SDNALKEVLLPASLKTITREAFKKNHLKQLEVASALSASHIAENALDDNDGDGDFONKVVVK 660  
QY 661 THNSYALADGEHFI VDPDKLSSTIVLEKILKIEGLDYSTLRQTQTQFRDMTTAGKA 720  
DB 661 THNSYALADGEHFI VDPDKLSSTIVLEKILKIEGLDYSTLRQTQTQFRDMTTAGKA 720  
QY 721 LLSKSNLRQGEKQFLQEAQFPLGRVLDLDAIAKAERKALVTKATQNGQLLERSINKAVL 780  
DB 721 LLSKSNLRQGEKQFLQEAQFPLGRVLDLDAIAKAERKALVTKATQNGQLLERSINKAVL 780  
QY 781 AYNSAISKANVKELEKELDLITGLVEGKGLAQATWVGYYLAKTLPPLPEYYIGLVNY 840  
DB 781 AYNSAISKANVKELEKELDLITGLVEGKGLAQATWVGYYLAKTLPPLPEYYIGLVNY 840  
QY 841 PDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDDNEGYPHALAVATLADYEGLDIKTILN 900  
DB 841 PDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDDNEGYPHALAVATLADYEGLDIKTILN 900  
QY 901 SKLSQTSIRQVPTAAVHRAGIFQAIQNAABAEQQLPKPOTHSEKSSSSSSANSKORGL 960  
DB 901 SKLSQTSIRQVPTAAVHRAGIFQAIQNAABAEQQLPKPOTHSEKSSSSSSANSKORGL 960  
QY 961 QSNPKTNRGRHSAILPRTGSKGSFVYGLVTSVALLSLITAIKKKY 1008  
DB 961 QSNPKTNRGRHSAILPRTGSKGSFVYGLVTSVALLSLITAIKKKY 1008

RESULT 2  
QSP1F7  
ID Q8P1F7 PRELIMINARY; PRT; 1008 AA.  
AC Q8P1F7;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Hypothetical protein spyM18\_0903.  
GN OrderedLocustNames=spyM18\_0903;  
OS Streptococcus pyogenes (serotype M18).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=186103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MGAS8232;  
RX MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;  
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
RA Sylva G.L., Sturdevant D.B., Ricklefs S.M., Porcella S.F.,  
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,  
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;  
RT "Genome sequence and comparative microarray analysis of serotype M18  
group A Streptococcus strains associated with acute rheumatic fever  
outbreaks.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).  
DR EMBL; AE010019; AAL9753.1; -  
DR GO; GO:0009986; C:cell surface; IEA.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR007093; LRR\_Tp.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
KW Complete proteome; Hypothetical protein.

SQ	SEQUENCE	1008 AA; 111620 MW; 89E79683FC2FF03B CRC64;
	Query Match	99.6%; Score 5058; DB 2; Length 1008;
	Best Local Similarity	99.6%; Pred. No. 2.3e-214;
	Matches 1004; Conservative	1; Mismatches 3; Indels 0; Gaps 0;
QY	1	MKKHLKTLVTLTTVVVTHNQEVSLVKEPILKQTQASSISGADYAESGSKLKINE 60
DB	1	MKKHLKTLVTLTTVVVTHNQEVSLVKEPILKQTQASSISGADYAESGSKLKINE 60
QY	61	TSGPVDVTVDLFSDKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120
DB	61	TSGPVDVTVDLFSDKRTTPEKIKONLAKPREQELKAVTENSEKQINSGSQLEQSK 120
QY	121	SLSLNTKVPSTSNWEICDFITKGNLTGLVLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
DB	121	SLSLNTKVPSTSNWEICDFITKGNLTGLVLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
QY	181	FAFTPKKTAIAEYTSRAGENGESQLDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
DB	181	FAFTPKKTAIAEYTSRAGENGESQLDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
QY	241	DAFVNKNIAEVLNPSLETISDYAFALHALKQIDLPNLKAIAGELAFDQITGKLSLP 300
DB	241	DAFVNKNIAEVLNPSLETISDYAFALHALKQIDLPNLKAIAGELAFDQITGKLSLP 300
QY	301	QRLMLAERAFKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360
DB	301	QRLMLAERAFKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360
QY	361	GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB	361	GDDHNNRVVLWTKSGKNPYGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
QY	421	GFNSKGLQKVRKNLEIPKQNGVTIIEIGDNAPRVDFQNTLRKYDLEEVKLPSTIR 480
DB	421	GFNSKGLQKVRKNLEIPKQNGVTIIEIGDNAPRVDFQNTLRKYDLEEVKLPSTIR 480
QY	481	KIDAFAFOSNNLKSFEASDDEIEKEGAFMNNRIETLEKDKLVIGDAAFHINHIYAIV 540
DB	481	KIDAFAFOSNNLKSFEASDDEIEKEGAFMNNRIETLEKDKLVIGDAAFHINHIYAIV 540
QY	541	LPBSVOEIGRSFRQNGANNLIFMGSVKVTLGEMAFNLNRLBHLDLSEKQKLTETPVOAF 600
DB	541	LPBSVOEIGRSFRQNGANNLIFMGSVKVTLGEMAFNLNRLBHLDLSEKQKLTETPVOAF 600
QY	601	SDNALKEVLLPASLTIREEAPKQHLKQLEVASALSHIAFNALDNDGDEQFQNKVVVK 660
DB	601	SDNALKEVLLPASLTIREEAPKQHLKQLEVASALSHIAFNALDNDGDEQFQNKVVVK 660
QY	661	THNSYALADGBHFIVDPDKLSSTIVDLKILIEGLDYSLRTTOTOPRDMTTAGKA 720
DB	661	THNSYALADGBHFIVDPDKLSSTIVDLKILIEGLDYSLRTTOTOPRDMTTAGKA 720
QY	721	LLSKNLRQGEKQKLOEAFQFLGRVDLDKATAEAKALVTKKATKNGQLLERSINKAVL 780
DB	721	LLSKNLRQGEKQKLOEAFQFLGRVDLDKATAEAKALVTKKATKNGQLLERSINKAVL 780
QY	781	AYNNSAIKANKVRLEKELDLTLGLVEGKPLAQATWQGVYLLKTPILPLPEYIYGLNYY 840
DB	781	AYNNSAIKANKVRLEKELDLTLGLVEGKPLAQATWQGVYLLKTPILPLPEYIYGLNYY 840
QY	841	PDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGVHALAVATLADYEGLDIKTILN 900
DB	841	PDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGVHALAVATLADYEGLDIKTILN 900
QY	901	SKLSQLTISRQVPTAAYTHRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKORGL 960
DB	901	SKLSQLTISRQVPTAAYTHRAGIFQAIQNAAAAEQQLPKPGTHSEKSSSESANSKORGL 960
QY	961	QSNPKTNRGRSAILPRTGSGKSFYVIGILGYTSVALLSLITAIKCKKY 1008
DB	961	QSNPKTNRGRSAILPRTGSGKSFYVIGILGYTSVALLSLITAIKCKKY 1008

RESULT 3

Q87817	PRELIMINARY;	PRT; 1008 AA.
ID	Q87817	
AC	Q87817;	
DT	01-JUN-2003 (TReMBLrel. 24, Created)	
DT	01-JUN-2003 (TReMBLrel. 24, Last sequence update)	
DT	01-MAR-2004 (TReMBLrel. 26, Last annotation update)	
DE	Hypothetical protein Spel285.	
GN	OrderedLocusNames=Spel285;	
OS	Streptococcus pyogenes (serotype M3).	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
OC	Streptococcus.	
OX	NCBI TaxID=198466;	
RN	[1]	SEQUENCE FROM N.A.
RP	SEQUENCE FROM N.A.	
RC	STRAIN=SSI-1;	
RX	MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;	
RA	Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,	
RA	Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,	
RA	Havashi H., Hattori M., Hamada S.;	
RT	"Genome sequence of an M3 strain of Streptococcus pyogenes reveals a	
RT	large-scale genomic rearrangement in invasive strains and new insights	
RT	into phage evolution."	
RL	Genome Res. 13:1042-1055 (2003).	
CC	-!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by	
CC	an amide bond (By similarity).	
CC	EMBL; AP005145; BAC64380.1; -.	
DR	GO; GO:0009986; C:cell surface; IEA.	
DR	GO; GO:0005618; C:cell wall; IEA.	
DR	InterPro; IPR001899; Gram_pos_anchor.	
DR	InterPro; IPR007093; LRR_Tp.	
DR	Pfam; PF00746; Gram_pos_anchor; 1.	
DR	TIGRPFAM; TIGR01167; LPATG_anchor; 1.	
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.	
KW	Cell wall; Hypothetical protein; Peptidoglycan-anchor.	
SQ	SEQUENCE 1008 AA; 111537 MW; 793C598CE094EB92 CRC64;	
	Query Match	99.2%; Score 5040; DB 2; Length 1008;
	Best Local Similarity	99.2%; Pred. No. 1.4e-213;
	Matches 1000; Conservative	3; Mismatches 5; Indels 0; Gaps 0;
QY	1	MKKHLKTLVTLTTVVVTHNQEVSLVKEPILKQTQASSISGADYAESGSKLKINE 60
DB	1	MKKHLKTLVTLTTVVVTHNQEVSLVKEPILKQTQASSISGADYAESGSKLKINE 60
QY	61	TSGPVDVTVDLFSDKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120
DB	61	TSGPVDVTVDLFSDKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120
QY	121	SLSLNTKVPSTSNWEICDFITKGNLTGLVLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
DB	121	SLSLNTKVPSTSNWEICDFITKGNLTGLVLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180
QY	181	FAFTPKKTAIAEYTSRAGENGESQLDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
DB	181	FAFTPKKTAIAEYTSRAGENGESQLDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240
QY	241	DAFVNKNIAEVLNPSLETISDYAFALHALKQIDLPNLKAIAGELAFDQITGKLSLP 300
DB	241	DAFVNKNIAEVLNPSLETISDYAFALHALKQIDLPNLKAIAGELAFDQITGKLSLP 300
QY	301	QRLMLAERAFKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360
DB	301	QRLMLAERAFKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360
QY	361	GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB	361	GDDHNNRVVLWTKSGKNPYGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
QY	421	GFNSKGLQKVRKNLEIPKQNGVTIIEIGDNAPRVDFQNTLRKYDLEEVKLPSTIR 480
DB	421	GFNSKGLQKVRKNLEIPKQNGVTIIEIGDNAPRVDFQNTLRKYDLEEVKLPSTIR 480

Db 421 GPSSKGLQKVKENKLEIPKQHNGVYTIETIGDNAFRNVDFQNKTLRKLYDLBEVKLPSTIR 480  
 Qy 481 KIGAFAPQSNLLKSPASDDLEIEIKEGAFPMNRIETLEBKOKLVITIGDAAFHINHIYAI 540  
 Db 481 KIGAFAPQSNLLKSPASDDLEIEIKEGAFPMNRIETLEBKOKLVITIGDAAFHINHIYAI 540  
 Qy 541 LPESVOETGRSAFRONGANNLIFMGSKVKTIGEMAPLSNRLEHLDLSEKQLTEIPVOAF 600  
 Db 541 LPESVOETGRSAFRONGANNLIFMGSKVKTIGEMAPLSNRLEHLDLSEKQLTEIPVOAF 600  
 Qy 601 SDNALKEVLLPASLKTIREEAFKQHLKQLEVASALSIAFNALDDNDGDEOFNKVVVK 660  
 Db 601 SDNALKEVLLPASLKTIREEAFKQHLKQLEVASALSIAFNALDDNDGDEOFNKVVVK 660  
 Qy 661 THNSYALADGSHFTVDPDKLSTSTIVDEKILKLEGLDYSTLRQTOTQFRDMTMTAGKA 720  
 Db 661 THNSYALADGSHFTVDPDKLSTSTIVDEKILKLEGLDYSTLRQTOTQFRDMTMTAGKA 720  
 Qy 721 LLSKSNLRQGEKQKQLEQAQFPLGRVLDLKAIAKAERKALVTKATNGQLLERSINKAVL 780  
 Db 721 LLSKSNLRQGEKQKQLEQAQFPLGRVLDLKAIAKAERKALVTKATNGQLLERSINKAVL 780  
 Qy 781 AYNSAIKANKVKRLEKELDLITGLVEGKGPLAQATMWQGVYLLKTPPLPEYIIGLVNY 840  
 Db 781 AYNSAIKANKVKRLEKELDLITGLVEGKGPLAQATMWQGVYLLKTPPLPEYIIGLVNY 840  
 Qy 841 FDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGHALAVATLADYEGLDIKTILN 900  
 Db 841 FDKSKLIYALDMSDTIGEGQKDAYGNPILNVDENEGHALAVATLADYEGLDIKTILN 900  
 Qy 901 SKLSQLTSTIRQVPTAAYHRAGIFQAIQNAABAEQQLPKPGTHSEKSSSESANSKORGL 960  
 Db 901 SKLSQLTSTIRQVPTAAYHRAGIFQAIQNAABAEQQLPKPGTHSEKSSSESANSKORGL 960  
 Qy 961 QSNPTNKGHSAILPRTGSGKSFVYGLGYTSVALLSLITAIKKKY 1008  
 Db 961 QSNPTNKGHSAILPRTGSGKSFVYGLGYTSVALLSLITAIKKKY 1008

RESULT 4  
 Q8K7X8  
 ID Q8K7X8 PRELIMINARY; PRT; 999 AA.  
 AC Q8K7X8;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Putative surface antigen.  
 GN OrderedLocusNames=SPV3 0569;  
 OS Streptococcus pyogenes (serotype M3).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=198466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MGAS315 / Serotype M3;  
 RX MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;  
 RA Beres S.B., Sylva G.B., Barbican K.D., Lei B., Hoff J.S.,  
 RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,  
 RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,  
 RA Schlievert P.N., Musser J.M.,  
 RT "Genome sequence of a serotype M3 strain of group A Streptococcus:  
 RT phage-encoded toxins, the high-virulence phenotype, and clone  
 RT emergence."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).  
 CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
 CC an amide bond (By similarity).  
 DR EMBL; AE014148; AA079176.1; -.  
 DR GO; GO:0009986; C:cell surface; IEA.  
 DR GO; GO:0005618; C:cell wall; IEA.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR007093; LRR\_Tp.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR TIGRFAMs; TIGR01167; LPTTG\_anchor; 1.

DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
 KW Cell wall; Complete proteome; Peptidoglycan-anchor.  
 SQ SEQUENCE 999 AA; 110518 MW; 39F9801882BAEF7C CRC64;  
 Query Match 98.3%; Score 4993; DB 2; Length 999;  
 Best Local Similarity 99.1%; Pred. No. 1.6e-211;  
 Matches 990; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 Qy 10 LTLTTSVVTNQNQVSLVKEPILKQTAQSSISGADYAESGSKLKINETSFPVDDTV 69  
 Db :  
 1 MTLTTSVVTNQNQVSLVKEPILKQTAQSSISGADYAESGSKLKINETSFPVDDTV 60  
 Qy 70 TDLFSDKRTTPEKIKDNLAKGPREQLKAVTENTSEKQITSGSQLESKESLSINKTVP 129  
 Db TDLFSDKRTTPEKIKDNLAKGPREQLKAVTENTSEKQITSGSQLESKESLSINKRVP 120  
 Qy 130 STSNWEICDFTTKGNTLVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVASFATPDKKT 189  
 Db STSNWEICDFTTKGNTLVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVASFATPDKKT 180  
 Qy 121 STSNWEICDFTTKGNTLVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVASFATPDKKT 180  
 Db :  
 190 AIAEYTSRAGENGISQLDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNV 249  
 Db AIAEYTSRAGENGISQLDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNV 240  
 Qy 250 AEVNLPSLETTISDYAFALHALKQIDLPDNLKAIGELAFPDNOITGKLSLPROLMRLAER 309  
 Db AEVNLPSLETTISDYAFALHALKQIDLPDNLKAIGELAFPDNOITGKLSLPROLMRLAER 300  
 Qy 310 AFKSNHKTTEFRGNLSKVIGEASFQNDLSQMLPDGLEKIESEAFTEGNGPDHNNRV 369  
 Db AFKSNHKTTEFRGNLSKVIGEASFQNDLSQMLPDGLEKIESEAFTEGNGPDHNNRV 360  
 Qy 301 AFKSNHKTTEFRGNLSKVIGEASFQNDLSQMLPDGLEKIESEAFTEGNGPDHNNRV 360  
 Db :  
 370 VLMTKSGKNPGLATENTYVNPDKSLMQESPEIDYTKWLEEDFTYQKNSVTFGSKGLQK 429  
 Db VLMTKSGKNPGLATENTYVNPDKSLMQESPEIDYTKWLEEDFTYQKNSVTFGSKGLQK 420  
 Qy 430 VKRNKNLEIPKHNGVITTEIGDNAFRNVDFQNKTLRKLYDLBEVKLPSTIRKIGAFAPQS 489  
 Db VKRNKNLEIPKHNGVITTEIGDNAFRNVDFQNKTLRKLYDLBEVKLPSTIRKIGAFAPQS 480  
 Qy 490 NNLKSPASDDLEIEIKEGAFPMNRIETLEBKOKLVITIGDAAFHINHIYAI 549  
 Db NNLKSPASDDLEIEIKEGAFPMNRIETLEBKOKLVITIGDAAFHINHIYAI 540  
 Qy 550 RSAPFRQGANLIFMGSKVKTIGEMAPLSNRLEHLDLSEKQLTEIPVOAFSDNALKEVL 609  
 Db RSAPFRQGANLIFMGSKVKTIGEMAPLSNRLEHLDLSEKQLTEIPVOAFSDNALKEVL 600  
 Qy 610 LPASLKTIREEAFKQHLKQLEVASALSIAFNALDDNDGDEOFNKVVVTHNSYALA 669  
 Db LPASLKTIREEAFKQHLKQLEVASALSIAFNALDDNDGDEOFNKVVVTHNSYALA 660  
 Qy 670 DGEHFI VDPDKLSTSTIVDEKILKLEGLDYSTLRQTOTQFRDMTTAGKALLSKSLNRQ 729  
 Db DGEHFI VDPDKLSTSTIVDEKILKLEGLDYSTLRQTOTQFRDMTTAGKALLSKSLNRQ 720  
 Qy 730 GEKQKFLQEAQFFLGRVLDLKAIAKAERKALVTKATNGQLLERSINKAVLAYNSAIKK 789  
 Db GEKQKFLQEAQFFLGRVLDLKAIAKAERKALVTKATNGQLLERSINKAVLAYNSAIKK 780  
 Qy 790 ANVRLSEKELDLITGLVEGKGPLAQATMWQGVYLLKTPPLPEYIIGLVNYFDSKGLIY 849  
 Db ANVRLSEKELDLITGLVEGKGPLAQATMWQGVYLLKTPPLPEYIIGLVNYFDSKGLIY 840  
 Qy 850 ALDMSDTIGEGQKDAYGNPILNVDENEGHALAVATLADYEGLDIKTILNSKLSQLTSTI 909  
 Db ALDMSDTIGEGQKDAYGNPILNVDENEGHALAVATLADYEGLDIKTILNSKLSQLTSTI 900  
 Qy 910 RQVPTAAYHRAGIFQAIQNAABAEQQLPKPGTHSEKSSSESANSKORGLQSNPKTNRG 969  
 Db RQVPTAAYHRAGIFQAIQNAABAEQQLPKPGTHSEKSSSESANSKORGLQSNPKTNRG 960  
 Qy 970 RBSAILPRTGSGKSFVYGLGYTSVALLSLITAIKKKY 1008

```
Db 961 RHAIPRTGSGSFYGLVTSVALLSIAAIAKKKY 999
|||||
RESULT 5
QBE1D6 PRELIMINARY; PRT; 1055 AA.
ID QBE1D6
AC QBE1D6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cell wall surface anchor family protein.
GN OrderedLocusNames=SAG0421;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=2222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tetelin H., Masignani V., Cielesiewicz M.J., Eisen J.A., Peterson S.N.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J.J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014211; AAM93327.1; -
DR TIGR; SAG0421; -
DR InterPro; IPR007093; LRR_Tp.
KW Complete proteome.
SQ SEQUENCE 1055 AA; 118129 MW; 2DD8B8715B17F74E CRC64;

Query Match 74.3%; Score 3776.5; DB 2; Length 1055;
Best Local Similarity 71.7%; Pred. No. 5.9e-158;
Matches 756; Conservative 105; Mismatches 140; Indels 53; Gaps 5;

QY 2 KKHXTVALTITVSVTHNQEVFSLVKEPIKQTAQSSISGADYAESGSKLKNET 61
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 KKHXTLALITVSVTHNQEVFSLVKEPIKQTAQSSISGADYAESGSKLKNET 61
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 SGFVDDTVTDLPSDKRTTPEKIKONLAKGPRQELKAVTENT-ESEKQITSGSQLEQSK 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 NSTVDETVDLPSDGNSSNNSKTESVSDPKQPKAKPEVTEASNSNDASKVEVPKQ 121
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 SISLNTKVPSTSNWEICDPIKNTLVGLSKGVEKLSQTHLVLPQAAADGTQLIOVAS 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 DTASKKETLETSTWAKDFVTRGDTLVGFSKGINKLSQTSHLVLPFSAADGTQLTQVAS 181
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 FAFTPDKKTATAYTSRAGENEISQDLDVGKEIINEGEVFNLSYLLKVTPTGKYGQ 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 FAFTPDKKTATAYTSRAGENEISQDLDVGKEIINEGEVFNLSYLLKVTPTGKYGQ 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 DAFVDMKNTAEVNLPSLETISDYAPAHALAKQIDLPNLKALGELAFDNDQITCKLSLP 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 242 DAFVDMKNTAEVNLPSLETISDYAPAHALAKQIDLPNLKALGELAFDNDQITCKLSLP 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 ROLMELAEAFKSNHIKTEPFGNSLVKIGEASFQNDLSQMLPDGLEKTESEFTGNP 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 302 RHLIKLAERAFKSNRIQTVEFLGSKLVKIGEASFQNDLRNMLPDGLEKTESEFTGNP 361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 361 GDHYNNRVLVTKGKPSGLATENTVYVDPKSLWQSSPEIDYTKWLEEDFTYKNSVT 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 362 GDEHYNNQVLETRTGQPHQLATENTVYVDPKSLWQSSPEIDYTKWLEEDFTYKNSVT 421
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 421 QFSNKGLOKVRNKNLEIPKQHGNTITEIGDNAPFNVDFOKTLRKYDLBEIKLPSTIR 480
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 422 QFSNKGLOKVRNKNLEIPKQHGNTITEIGDNAPFNVDFOKTLRKYDLBEIKLPSTIR 481
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 481 KIGAFAFOSNNLKSPFASDDLEETKEGAFMNRLETLELKDKLVITIGDAAFHINHYAIV 540
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 482 KIGAFAFOSNNLKSPFASDDLEETKEGAFMNRLETLELKDKLVITIGDAAFHINHYAIV 541
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 541 LPESVQEIIGRSAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDLSBQKQLTEIPVQAF 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 542 LPESVQEIIGRSAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDLSBQKQLTEIPVQAF 601
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 601 SDNALKEVLLPASLKTITREAFKKNHKLQEVASALSALHAFNALDDNDGDFQDKVVKV 660
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 602 SDNALKEVLLPASLKTITREAFKKNHKLQEVASALSALHAFNALDDNDGDFQDKVVKV 661
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 661 THNSYALADGEHPITVDPKLSSTITVDLEKILKIEGLDYSLRQTTOTQFRDMMTAGKA 720
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 662 THNSYALADGEHPITVDPKLSSTITVDLEKILKIEGLDYSLRQTTOTQFRDMMTAGKA 721
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 721 LLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTCKATKNGQLLERSINKAVL 780
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 722 LLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTCKATKNGQLLERSINKAVL 781
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 781 AYNNSAIKANVKKLEKELDLITGLVEGKGLAQATWQGVYLLKTPLPPEYIYGLNVY 840
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 782 AYNNSAIKANVKKLEKELDLITGLVEGKGLAQATWQGVYLLKTPLPPEYIYGLNVY 841
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDENEGHALAVATLADYEGLYIKDILN 900
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 842 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDENEGHALAVATLADYEGLYIKDILN 901
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 901 SKLSQLTIRQVPTAAVHRAGIFQAIQNAAEAEQLPKPGTHSEKSSSSSANSKDRGL 960
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 902 SSLLDKIKAIRQIPLAKYHRLGIFQAIRNAEADELPLK-----TPKGYLVNEVPYKKQV 957
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 961 QSNPK-----TN-----RGRHSA 973
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 958 EKNLKPVDYKTFIFNKALPNEKVDRAAKGHINAEITNSNNSVAVTPIRSEQQLKSDV 1017
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 974 ILPRGSGSFYGLVTSVALLSLTAIKKK 1007
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1018 NLPTSSKNNFIYELGVSLCLFLVTAGKKGK 1051
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 6
QBE6V3 PRELIMINARY; PRT; 1055 AA.
ID QBE6V3
AC QBE6V3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein gbe0456.
GN OrderedLocusNames=gbe0456;
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=2242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalloui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766845; CAD46100.1; -
DR SacalList; gbe0456; -
DR InterPro; IPR007093; LRR_Tp
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1055 AA; 117976 MW; 9DD56FB9AD171BD8 CRC64;

Query Match 74.2%; Score 3771.5; DB 2; Length 1055;
```

Best Local Similarity 72.0%; Pred. No. 9.8e-158;		Matches 756; Conservative 103; Mismatches 146; Indels 45; Gaps 5;	
Qy	2	KKHLKTVLTLTVSVTHNQVPSLVKEPIKQTKQASSISGADYAESGSKSLKINET	61
Db	3	KKHLKTLALATLTVSVTVSQEVGLREESVKQEQTQSA-SEDDWFEEEDNERKTNVSKG	61
Qy	62	SGPVDDTVTDLPDRKTRTPEKIKONLAKGPRQBELKAVTENT-ESEKQITSGSQLEBQKE	120
Db	62	NSTVDETVSDLPFSDGNSNNSKSTESVSDPKQVPKAKPEVTQEASNSNDASKVEVPKQ	121
Qy	121	SLSLNTKVPSTSNWEICDFITKGNLTVLGSLKSGVEKLSQTDHLVLPQAAADGTOLIOVAS	180
Db	122	DTASKKETIETSTWEAKQFVTRGDTLVGFSGKINGKLSQTSHLVLPQAAADGTQTQVAS	181
Qy	181	FAFTPDKKTAAIAYTSRAGENGESIQDVGKEIINEGEVFNLSYLLKKVTIPTGYKHIQ	240
Db	182	FAFTPDKKTAAIAYTSRAGENGKPSRLDIDQKEIIDEGEIFNAYQLTKLTPNGYKSIQ	241
Qy	241	DAFVNKQIAEAVNLPSLESLETSIDYAFALHALKQIDLPNKLKAIGELAFDNOITQKLSLP	300
Db	242	DAFVNKQIAEAVNLPSLESLETSIDYAFALHMSLKQVLPNKLKVGELAFDNOIGGKLYLP	301
Qy	301	ROLMLAERAPKSNHIIKTEPRGNLSKVIGEASFODNDLSQMLPDGLEKTESRAFTGNP	360
Db	302	RHLIKLAERAPKSNRIQIVFELGSKLVIGEASFODNLSNVMPLDGLEKTESRAFTGNP	361
Qy	361	GDDHNNRVLWTKSGKQPSGLATENTYVNDKSLWQESPEIDTYKMLEEDFTYQKNSVT	420
Db	362	GDEHNNQVLRTRTGQNPQLATENTYVNDKSLWRATPDMDYTKMLEEDFTYQKNSVT	421
Qy	421	GFSNKGLOKQKQKQLEIPKQNGVITTEIGDGNAPRVDPQNTURKYDLEBEVKLPSTIR	480
Db	422	GFSNKGLOKQKQKQLEIPKQNGVITTEIGDGNAPRVDPQNTURKYDLEBEVKLPSTIR	481
Qy	481	KIGAFAPQSNLKSPEASDDLEIEKEGAPMNNRIETLEKOKLVITGDAAFHNNHIAIV	540
Db	482	KIGAFAPQSNLKSPEASDDLEIEKEGAPMNNRIETLEKOKLVITGDAAFHNNHIAIV	541
Qy	541	LPESVOEIGRSAPRONGANNLIFMGSVKVTLGEMAFSLNRLEHLDLSBQKOLTEIPVQAF	600
Db	542	LPESVOEIGRSAPRONGALHLWFIKGVKVTIGEMAFSLNKLESVNLSEKQLKTIEVQAF	601
Qy	601	SDNALKEVLLPASLKTIREBAPKQHLKQLEVASALSHIAFNALDDNDGDSQFQDNKVVK	660
Db	602	SDNALSEVLLPPLQOTIREBAPKRNHLKEVKGSTLSQIAFNAPFQNDGDKRFGKKVYVR	661
Qy	661	THNSYALADGHRFTVDPKLSSTVLEKILKLTIEGLDYSTLRQTTOTQPRDMTTAGKA	720
Db	662	THNSHMLADGHRFTVDPKLSSTVWDLKVLKLTIEGLDYSTLRQTTOTQPREMTTAGKA	721
Qy	721	LLSKNLKQGEKQKQFLOBAQFPLGRVDLDKAKAEKALVTKATKNGQLLERSINKAVL	780
Db	722	LLSKNLKQGEKQKQFLOBAQFPLGRVDLDKAKAEKALVTKATKNGHLLERSINKAVL	781
Qy	781	AYNSAIKANKYKRLKELDLTLTGVEKGPLAQATWQGVYLLKTPLPLEYYIIGLVNY	840
Db	782	AYNSAIKANKYKRLKELDLTLTGVEKGPLAQATWQGVYLLKTPLPLEYYIIGLVNY	841
Qy	841	PKSGLIYALDMSDTTIGEGQKDAVGNPLNVDEONEGVHALAVATLADYEGLDIKTILN	900
Db	842	PKSGLIYALDMSDTTIGEGQKDAVGNPLNVDEONEGVHALAVATLADYEGLYIKDILN	901
Qy	901	SKLSQTSIRQVPTAAYHRAGIFQIONAAAEAEQLPKP-----GTHSEKSSSSSESANS	955
Db	902	SSLDIKAIQIPLAKYHRLGFPQAIRNAAAEADRLIPKTPKGYLNTVPNYKQKQVEKNS	961
Qy	956	KD-----RGLQSNPKTN-----RGRHSAIILPR	977
Db	962	KPVDTKPIFDKALPNEKVGDRVAKGHNNIAETNNYVAVTPIRSEQQHLKSSQSDVNLQ	1021
Qy	978	TGSGKSFVYGILGYTSVALLSLITAIKKKK 1007	

Db

1022

TSSKNFIYEILGYVSLCLLFLVLTAGKKG

1051

RESULT 7

O68831

PRELIMINARY;

PRT;

1081

AA.

AC

O68831,

01-AUG-1998

(TrEMBLrel. 07, Created)

DT

01-AUG-1998

(TrEMBLrel. 07, Last sequence update)

DT

01-OCT-2003

(TrEMBLrel. 25, Last annotation update)

DE

Surface antigen BspA.

GN

Name=BspA;

OS

Bacteroides forsythus.

OC

Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;

OC

Porphyromonadaceae; Tannerella.

OX

NCBI\_TaxID=28112;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

STRAIN=ATCC43037;

RX

MEDLINE=99043895; PubMed=9826345;

RA

Sharma A., Sojar H.T., Glurich I., Honma K., Kuramitsu H.K.,

RA

Genco R.J.;

RT

"Cloning, expression, and sequencing of a cell surface antigen containing a leucine-rich repeat motif from Bacteroides forsythus ATCC 43037.";

RL

Infect. Immun. 66:5703-5710(1998).

DR

EMBL; AF054892; AAC82625.1; -.

DR

PIR; T31094; T31094.

DR

InterPro; IPR003343; Big\_2.

DR

InterPro; IPR007093; LRR\_Tp.

DR

Pfam; PF02368; Big\_2; 1.

DR

SMART; SM00635; BID\_2; 1.

SQ

SEQUENCE 1081 AA; 113907 MW; 65E36FA0ACBD5C CRC64;

Query Match

6.5%;

Score 331.5;

DB 2;

Length 1081;

Best Local Similarity

25.3%;

Pred. No. 2.4e-06;

Matches 147;

Conservative 83;

Mismatches 206;

Indels 144;

Gaps 23;

Qy

110

TSGQLSQSKESLSLN-----KTVFSTSNWEICDFITKGNLTVLGSLKSGVEKLSQTDH

162

Db

19

TLGATAQNSGTTGLNWSVDSGTKTLAITGTGAMPDF---NNASEIPMHSLSQSKIQT--

72

Qy

163

LVLPQAAADGTQLQVASFAT-----PDKKTAAEYTSRAGENGEISQLDVGKE

213

Db

73

-----VTIGDVTSVGNNAFSDCALTSVTLPSNLTAIGDHAFK-GCSGLTSITIPNSVT

125

Qy

214

IINEGEVFNLSYLLKKVTIPTGYKHIGQDAFVDNKNIAEAVNLPESETISDYAFALHA-LK

272

Db

126

TIGEWAPKGCGLKSLITLPSNLTAIGQSLSGCTGLTSITIPNSVTTIGEWAFGCCSLT

185

Qy

273

QIDLPDLNKAIGELAFDNOITGKLSLPQMLRLAERAPKS-NHIKTIETFGNSLKVIGE

331

Db

186

SITFPNSLTAIGESAFYCGGALTSTLTPDALTTIGESAFKGCGLKSLTF-PNSLTITGE

244

Qy

332

ASPOD-NDLSQMLPDGLEKIESAFTGNPDHNNRVNLWTKSGKNPSGLATENTYVN

390

Db

245

SAFYDCGALTSITLTPDALTTIGRSAFYG-----CSGL-----

276

Qy

391

PKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGKLVKRNKNLEIPKQHNGVTITEI

450

Db

277

-----KSITFPN-----SLTTI

288

Qy

451

GDNAFRNVDFONKTLRKVDLEEVKLPSTIRKIGAFAPQS-NNLKSFASDDELEIKGAF

509

Db

289

GESAFYNGC-----SLTSITIPNSVTTIGRSAFYCGGSLKSLITLPGLTITIERAF

339

Qy

510

MNNRIET-LELKOKLVITGDAAFH-INHIYAVLPESVQEIGRSAPRONGANNLIFWGSK

567

Db

340

YNCGVLSITIPNSVATIGESAFYCGGSLKSLITLPGLTITIEWGAFYNGCALTSITIPNS

399

Qy

568

VKTIGENAFSL-NRLEHLDLSEKQLTEIPVQ--AFSDNALKEVLL--PASLKTIREBAP

622

Db

400

VSTIGESAFYCGGALKQDVTVAWD---TPIDIQDVFRRLTSLGIRLHVPAKKTIVYE---

453



```

QY 623 KQHHLKOLEVASALSIAFNALDND-GDEQFDNKVVVKT 661
DB 454 AKDVKWE-----FNIVEDDDFGGLQWNYDAATKI 482

RESULT 8
Q8TI64
ID Q8TI64 PRELIMINARY; PRT; 1995 AA.
AC Q8TI64;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Cell surface protein.
GN OrderedLocusNames=MA4292;
OS Methanococcus acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Gargan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McSwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542 (2002).
DR EMBL; AE011143; AA007636.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR010983; EF_Hand_like.
DR InterPro; IPR007093; LRR_Tp.
DR InterPro; IPR006061; PKD.
DR Pfam; PF00801; PKD; 6.
DR SMART; SM00089; PKD; 6.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS50093; PKD; 6.
KW Complete proteome.
SQ SEQUENCE 1995 AA; 208421 MW; 031C7F16092BE035 CRC64;

Query Match
Best Local Similarity 6.2%; Score 317; DB 2; Length 1995;
Matches 177; Conservative 112; Mismatches 289; Indels 190; Gaps 36;

QY 14 TVSVWTHNDEVFSLVKEPILKQFASSTSGADYA-----ESSGSKLKI 58
DB 466 TVSLTVSN-----TGGSSSTETKADYIYAALPDGPFADSVDTVSGAPLTV 511
QY 59 N-----ETSGPV-----DDVTDLFSKRTTPEKIKDNLAKGPREQLKATENTESK 107
DB 512 QFTDLTNGFVSWANDFSDGIDSYEQNPSTYNAAGNYTVTMTATNL--VASDTVTRV 569
QY 108 QITSGSLEQ-----SKESLSLNKTPVTSNWEICD----- 138
DB 570 NLTSVENLPMADFSADPVATLIGCSVQFTDLSNSTPSQWQDFNNDGIDTSMQNPST 629
QY 139 FITKGNLTVLGSKS-----GVEKLSQTDHLVLPQADGTQLIQVASFPTPD-KKTAIAFY 194
DB 630 YTTAGTYTNLTVSNPAGSGDEKTDYIVVKEQAS-----HASDFTYSDGSSITITEY 683
QY 195 TSPAGENG-EISQLDVDGKEI--INEGEVFNSSLKKVTIPGYKHIGQDAFVKNKNAIE 251
DB 684 T---GSDGVIIPAEIEGLPVTITIGASAFYGCAPTPTVTPNSVTTIGDSAFQDCSALT 740

```

```
Db 555 VVKSIFYAFSSSSQLHSEIVLCSGVTKIEGYAFNPFSEITEITIPA--TVTDIDTNAP 612
Qy 78 TPEKIKD-NLAKGPREBELKAVTE-----NTESEKQITSG-----SOLEQSKESL 122
Db 613 ANCTSLKTVFEEKG-----SVTIGNAFYCNQITELRILSGDINFKCSLVQPNKIL 665
Qy 123 SLNKTVPSTNWEICDFITKNT-LVGLSKSGVEKL-----SQTDLHLVPSQA- 169
Db 666 SYLGLREVTS---ICSNAPSGCTQTLGDISHCKKLTSHAAAFNCSTSLVLEKLPASLE 722
Qy 170 -----ADGTQLIQVASFAPTPDKKTAIAEYTSRAGENG-----E 203
Db 723 FIDNSAFLSCSIAELDIAHCENLKSISTSFENCSNLKTVKLPKSIETIENRAPHYCYQ 782
Qy 204 ISQLDVD--GREIINEGVFNS-YLLKKVTIPTGYKHHGQDAFVQDNKNAEVLNPLET 260
Db 783 IEQVDSLPSYKLTIEDSVFSGCSALRSITFPDLERIGNSAFNVCYALTAVQPPKNLEQ 842
Qy 261 ISDYAFAH-LALKQIDLPNLKAIGELAFPDNQITGKLSLPLQMLRAERAFKS----- 313
Db 843 IGDSAFSNCYALTAVOLPNNLKRIKNYGFSDCSALTAVOLPNNLLEIGRSFYNCSAITG 902
Qy 314 ----NHKITE-----FRG-----NSLKVIGBASFOD-NDLSQLMLPDGLEKIE 352
Db 903 ITFGQKLTIQDNAFEKCRGIRGALDMRACNTLIEGGSAPFLDCSAITEIKLPASLQTLK 962
Qy 353 SEAP-----TQNGDDHYNRNVLTWTKSGKNPSGLATENTVYNPD- 392
Db 963 KRSPANCISLQKVLNGAIVTTTGTSGS-----FLKCKRSIKELTILSGKVTFKVDM 1015
Qy 393 -----KSL-----WQESPEIDYTKLE-----EDPTYQKNSVTGPNKGLQKVRKN 435
Db 1016 LADSYKTISSVVGVGDEVTIGSTAFSECLGKVLGDLG-DCDKLTSIGVSAPSGCSRLLG 1074
Qy 436 LEIPQHGQVTTITGDNAPRN-----VDPQ--NKTLRK-----YDLERVKLPST 478
Db 1075 VLPK-----NLTKIDSSAFSCIGSLVDMSCAKLVLSGSSAFYGCNSISIDIKFPKN 1129
Qy 479 IRKIGAPQOS--NNLKSGFEAS--DDEBEIKGAPMNN-RIETLEKDKLVITIGDAF-H 532
Db 1130 IKTIGSSCTNCTNYIAGSIDMSGCTELOKIENSIFSSDKKIQTAKLPNNITEIADSAFLK 1189
Qy 533 INHIAIVLPSGVQIGRSAPFQ-----NGAN-----559
Db 1190 CSGLEITELPESLTKIGNKAFGECTALASVEFGKGDILMATGNDASFYNDRDITKLSVLK 1249
Qy 560 -NLIFWGS-----KVKTL-----GEMAFLS-NRLE-HLDLSEQKOLTEIPVQA 599
Db 1250 GNIIFKCSDVFPHSYQKIKTLILGEGITGIGASAFSSCNGIEGTNLGNCRALIAIGNQA 1309
Qy 600 P-SBNALKEVLLPASLKTIREAPFK-----NHLKOLEVASALSHAFNALDNDGDEPFD 654
Db 1310 FNSCYGITGVQVPASLQTLIGELAFNGCNGINTMDLSPCTSMQSI-----GKSAPY 1360
Qy 655 NKVVVKTTHNSVALADGEHPIVDPKLST-----IVDLEKILK-----693
Db 1361 NCTKI-----TGKVPQOSLTSIGSTAFYQCVALEGMVDSLSCDLKIIDVGLFQG 1410
Qy 694 --LIEGLDYSTLRQTQTQFQDMTTAGKALLSKSLNRQCEKQKFLQEAQFPLGRVDL-- 749
Db 1411 CGLISGLKLP-----KTIITIGASAF--ENCKK-----IDAVDLSVC 1445
Qy 750 -----KAIKAKBALVTKATKNGQLLERSINKAVLA--YNNSAIKKANVGRLEKEL 799
Db 1446 ADLATTIGVQAFKSCAEAVITLPGKN-----SVGTAAFGNYSNDSYCKKVRIPR-GADG 1498
Qy 800 DLTGLGVEKG 810
Db 1499 DSLKGRVTSSG 1509
```

RESULT 10

```
Q8TI67
ID Q8TI67 PRELIMINARY; PRT; 1734 AA.
AC Q8TI67;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cell surface protein.
GN OrderedLocusNames=M4289;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1];
SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21922760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE011141; AAM07633.1; -.
DR GO; GO:0005509; P:calcium ion binding; IEA.
DR InterPro; IPR002048; EF-Hand.
DR InterPro; IPR010983; EF-Hand-like.
DR InterPro; IPR007093; LRR_Tp.
DR InterPro; IPR000601; PKD.
DR Pfam; PF00036; eFhand; 1.
DR Pfam; PF00801; PKD; 3.
DR SMART; SM00089; PKD; 3.
DR PROSITE; PS00018; EF HAND; UNKNOWN_1.
DR PROSITE; PS00093; PKD; 3.
DR Complete proteome.
SQ SEQUENCE 1734 AA; 181176 MW; AFA3EC48D2ADE877 CRC64;
Query Match 5.2%; Score 263; DB 2; Length 1734;
Best Local Similarity 22.0%; Pred. No. 0.0049;
Matches 167; Conservative 110; Mismatches 312; Indels 170; Gaps 35;
Qy 37 QASSISGADYAESGSKLKNETSGPVDVTVDLPSDKRTP-----80
Db 323 QSSVLSGA--VDLDNIYVRIVDVPGTG-----FKDSLGNFIYDAWLWTGTTGGFDLD 374
Qy 81 -----EKKONLAKGPREQELKAVTENTSESKQITSGSQLESLSLNTKVPST 132
Db 375 AVGVINSQKASDGSFASPTSGDTPLTQVFTDLNDGTTSMAW-----DPDNGDNDVSTE 428
Qy 133 NWEICDFITKNTVLVGLS---KSGVEKLSQTDHLVLPQAADGTQLIQVASFAT-PDKK 188
Db 429 QNPMTHTPTGYAVRLNVTVYSEGSDWKVTAYITVNTPAAN-----YTYTSGSS 480
Qy 189 TAIAEYTSRAGE---NGEISOLDVDGKEIINEGVFNSYL-LKKVTIPTGYKHHGQDAFV 244
Db 481 VTITGYSGPGGDVITPGEIEGLPVT-----IDDSVFMGYSNMTSVTLPDNVTGIGNNAFQ 536
Qy 245 DNKNAEVLNPLESETTSDYAFAHALKQIDLPNLKAIGELAFPD-----NQIT---GKL 297
Db 537 GCTNLTSVTFGDSITEIGSSAFQSTGLTSVVLPSVTSMGSSVFQDCTGLTSVIGNQTS 596
Qy 298 SLPQLMRLAERAFKSNHKTIEFRGNSLKVIGBASFOD-NDLSQLMLPDGLEKIESEAP 356
Db 597 SIPDKTF-----YRCDHLTSITI-SDSVTSIGKYAFRDCNTLTSVVLPSDTSIGAYAF 649
Qy 357 TG-----NPGDDHYNR---VVLWTKSGKNPSGLATENTY-----VNPDK 393
```

```

Db 650 EGCTALTSITIPDSVTSIDRYAFRGCTALTSATIGSGVTGIGVGFYQCTNMASTVTPD- 708
Qy 394 SLWQESPEI--DYTKWLEDFYQK--NSVTGFSNKGLOKVRKNKLEIPKQHGVTITE 449
Db 709 ----NVTIGNDAFWHC--TNLTSMTIGNVTTIGSFAFGCTGLASVTIPD-----SVTP 758
Qy 450 IGNAFRNVDYFQNTLRKYDLBEVPLPSTIRKIGAFQFOS--NNLKSFEASDDLEBIKEGA 508
Db 759 INDFAFGCT-----NLTSITLPNSVISIGNFVDFDCSALTSVTFSGSLTSGSNV 809
Qy 509 FNN-----NRIETLELXDLVTTGDAAFHINHIAIVL-----PESVOETGRSAP 553
Db 810 FOSCTALNAI--NVDANNVTYTSIDGIVYNNKDVAVVLPFSKAGSIVIPDSVTSISSYAF 868
Qy 554 RQNGANNLIPMGSKVKVTIGEMAFPLSNRLEHL--DLSEQKQLTEIPVQAFSD----- 602
Db 869 YGCAGLTSVTIGNSVTITGQAP--NGCTALTSITIPDRVTEIGICAFODCSALTSATPG 926
Qy 603 -----NA-----LKEVLLPASLKTIREAFKK--NHLKQLEVASALSIAFNALDDN 647
Db 927 TGLITIGENAFNGCTNLTSAMIPNNVTITIGSAFNGCTNLTSVMIGSGVTSIGANAKIAP 986
Qy 648 DGDEQPDNKKVVKTHNSVALADGEHFIVDPDKLSSTIV 686
Db 987 HGCTAL--TEVSVDAINTAYSSIDG-----VYDKAGTTLV 1020

RESULT 11
Q7PG02
ID Q7PG02 PRELIMINARY; PRT; 1066 AA.
AC Q7PG02;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000023103 (Fragment).
GN Name=ENSANG00000019883;
OS Anopheles gambiae str. PRST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxId=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAA01008846; EAA45120.1; -.
FT NON_TER 1 1
FT NON_TER 1066 1066
SQ SEQUENCE 1066 AA; 122246 MW; 9DE10C52881P98BA CRC64;

Query Match 4.9%; Score 247; DB 2; Length 1066;
Best Local Similarity 22.1%; Pred. No. 0.012;
Matches 205; Conservative 147; Mismatches 350; Indels 226; Gaps 41;

Qy 55 KUKINTSGPVDVTDVDFSDKRTTPEKIDNLAQPREQELKAVTENTSEKQITSGS- 113
Db 242 KTSMBESNAGLQSLDALEHTTKQLADNLEQEMAK---TADLNRLTEELEAEKQLMQEL 298
Qy 114 --QLEQSKESLSLNTKVTPTSNWEICDPITKGNLTVGLSKSGVEKLSQTDHLVLSQAAD 171
Db 299 ETQKQLMQELEVAKTSMSNNAL-----QRSUKAIEQT-----KMQLAD 339
Qy 172 GTQLIQVAFAPFPDKKTAIAEYTSRAGENGETSQLDVQKGTINGEVFNLSLLKKVTI 231
Db 340 NLE-----KEIAKTADLNRRTEELEVEKKQLTQLEVAKT----- 374
Qy 232 PTGYKHIGODAFVNDKNIATVNLPSLETISDYAFHAHLAKQIDLPDNL--KAIGELAFPD 290
Db 375 -----SKSENIAE--LQRSLEAIEQ-----TKQLADNLEKBIAKTADL- 411

```

```

Qy 291 NOITGKLSL--PROQLMRLAERAFKSNHIIKTIEFRGNSLKVIGEASFQDNDLSQMLPDGLE 349
Db 412 NRTEKLEVEKKQLMQELEVA--KTSMSENIAELQRSLEAI-----EQTKLQPADNLE 462
Qy 350 KIESAFTGNPGDDHNNRVVWLTKSGKVPGLATENTYVNPDKSLWQESPEIDYTKWLE 409
Db 463 KEIAKTADLNRRTEELEVEKKQLTQLEVAKSSGSENI-----AELQSLKAIEQTKLQL 517
Qy 410 EDFTYQKNSVTGFSNKGLOKVRKNK-----NLEIPKQHGVTITE-----I 450
Db 518 ADNLEKEIAKTADLNRRTEKLEVEKKQLMQELEVAKTSMSENIAELQSLKAIEQTKLQL 577
Qy 451 GDNAPRNV---DFQNTLRKYDLBEVPLPSTIRKIGAFQFOSNN-----LKSF 495
Db 578 ADNLEKEIAKTADLNRRTE--EKLEVEKKQLMQEL--EVAKTSMSENNAELQRSKALTELTKL 635
Qy 496 BASDDLE--EIKGAPMNNRIETLEKDKLVITGDAAFHINHIAIVLPESVOEIGRS--A 552
Db 636 QLADNLEKEIAKTADLNRRTEELEVEKKQL-----LQLEVAKTSMSENIAELQRSKLA 689
Qy 553 PRONG---ANNLI PMGSKVKVTIGEMAFPLSNRLEHLDLSEQKQLT--EIPVQAFS----- 601
Db 690 IEQTKLQLADNL-----EKSIKTAELNRRTEELEVE--EKKQLTQLEVAKSSSENIAT 742
Qy 602 -----DNALKEVLLPASL--KTIREEAFKKNHLKQLEVASALSIAFNAL 643
Db 743 LQRSLEAIEQTKLQLADNLEKEIAKTADLNRRTEELEVEKKQLMQELEVAKTSKSEIAT 802
Qy 644 LDDN--DGDEQFD-----NKVVVVKTHNSVALADGE-----HFIVDPDKLSSTIV 687
Db 803 LQOSLEAIEQTKLQLANNLEKEIAKTADLNRRTEELEAEKKQLMQELEVAKTSMSENNAE 862
Qy 688 LEKILKLEGLDYSLRTQTTQTFQDMWTAGKALLSKSNLR-----QGEKQKFLQEAQFPL 743
Db 863 LQRSKLAIE-----QTKQLADNLEKSIKTAELNRRTEELEVEKKQLTQBLE--V 911
Qy 744 GRVLDLDAKIAKAEKALVTKKATKNGQL---LERSINKAVLAYNNSAIKKNVVKLEKELD 800
Db 912 AKSSGSENIABEQSLKAIEQTKL--QLADNLEKEITKTADLNRRTEELEVEKKQLTQBLE 970
Qy 801 LLTGLVEGKGPLAQATMVQGVYLLKATPLPEYIYGLNVYFDKSGKL---IYALDMSDTI 857
Db 971 DM-----KTTNALSQSLVLE--LTQTQITNSLELVLVKIGDLQHLKAMDEAKTL 1020
Qy 858 GEGQKDAYGNPIL-----NVDEQNE 877
Db 1021 AEKLGDSERSEMLAKISLLQQLNEERE 1048

RESULT 12
Q8TI71
ID Q8TI71 PRELIMINARY; PRT; 1743 AA.
AC Q8TI71;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cell surface protein.
GN OrderedLocustNames=MA4285;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxId=22114;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G. Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Ainoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,
RA Linton L., McSwain P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grzame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,

```



QY 289 -FDNQITGKSLPQRLMLRAERAFKSNHIIKTIEFRGNSLKVIGEASFQNDLSQMLPDPG 347  
Db 4033 AIDNAVAGKLEINDSLTNEEQAYVDLIN-----NEADNAQKIAEATTPPE 4080  
QY 348 LEKIESEATGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDYTKW 407  
Db 4081 VTRAQEE-----GVKNHININVPPTSPAKDAANAIAI-----DQALNKKKDEINNATN 4127  
QY 408 L-----EEDFTYQKNSVTGFSNKGLOKVKKNKLEI-----PKOHNGVTITEIGDAFNVR 458  
Db 4128 ISSEKTDLIKQATEAANTAKONINNATTNSEVETAQVDGEKAIADVTVPGLSDIKKESI 4187  
QY 459 DPQNTKL-RKYDL-----EVLKPSITIRKIGAFAPQNNLKSFEASDLDEIKEG 507  
Db 4188 DLINKALNEKQDEINNASLSQDEKQELIDQAKKIAEAI-----NEINNAQTNDKAEADT 4245  
QY 508 AFWNNRIETLEKDKLVITIGDAAPHINHIYAIVLPESVOEIGRSAPRQNGANNLPMGSK 567  
Db 4246 GVKN-----IENVSIP-----STEDAKGNATQAIIDALNSKKNEI-----NNASNL----- 4286  
QY 568 VKTLAGEMAFSLRLEHLDLSEKQKLTETPQAFSDNALKEVLLPASLKTIREEAFKKNHL 627  
Db 4287 --TDSEKTDLIN-----QATEIA-----NAKDAINSATTNTAVEAAYKG-- 4325  
QY 628 KOLEVASALSHIAFNALDNDGDEQPDNKKVVKTHNSVALADGEHFIYDPDKLSTIYD 687  
Db 4326 ----VADINNIHFTNLDS-----KKAANS-AIEDALNTKKDEINNASLSD 4367  
QY 688 LEKILKLEGLDYSLRQTOTQFRDMMTAGKALLSKSNLRO-----GEKQKFLQE----- 738  
Db 4369 SEK-AKLI-----NQATEIANAARAAAINNATTNSAVTAAENKGIEDIANINV 4413  
QY 739 ---AOPFLGRVLDKAIKAERKALVTK---KATNGOLLERSINKAVLAYNSAIIKAN 791  
Db 4414 PSIAETKQAIIDAIQOVQAKNSQTEAKNSLGADEQKNLIDQ-VNKIA-----QDAILKLN 4468  
QY 792 -----VKRLEKELDLTL-----VEGKGPQAQATMVGVYLLKTLPLP 830  
Db 4469 DPATTNTEVITTRDKAIQITNLFIPTLDSVQKQAQEAINSQAQETKIDEI----- 4519  
QY 831 PRYITGLNVYFKSGKLI-----YALDMSDTTGEGOKDAYGNPILNVD-----EDNEG 879  
Db 4520 -----NKADNLTDQMKQNLTDQVDQADKATKAIINNAQTNDVKEAEIEGLEDI 4568  
QY 880 HALAVATLADYEGLDIKTI-----LNSK-----LSQTSIRQVPTAAVHR 919  
Db 4569 DSIKVPSLVEKODAIKEINDALKKKTDEINADLDQKQDELSQITDI-----ATETK 4623  
QY 920 AGIFOAIQNAABAE-----QLLPKPGTHSEKSSSESANSKDRGLQSN---PKTNR 968  
Db 4624 TKVFNATTNAEVDAAEAGIKAEAVKIIPARTADNSNT--ESESKEQVITNSVQPKRNA 4681  
QY 969 GRH-----SAILPRTGSK-----GSFYGILGYTSVALLSLITAIIKKK 1007  
Db 4682 VHHKNGTPVKNKATLPQTGKDKNSNLTLAGAALLGLAGVFS--LFLGLGDKRKNK 4734

## RESULT 14

Q81J57 PRELIMINARY; PRT; 1139 AA.  
AC Q81J57;  
DT 01-MAR-2003 (TReMBLrel. 23, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE QF122 antigen.  
GN ORFNames=PF10\_0115;  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,

RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Vaidya A.B.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrell B.,  
RA "Genome sequence of the human malaria parasite Plasmodium  
falciparum".  
RL Nature 419:498-511 (2002).  
CC -!- SIMILARITY: Contains 2 KH domains.  
DR EMBL; AE014830; AAN35313.1; -.  
DR HSSP; Q9UNW9; 1DTJ.  
DR GO; GO:0003676; P:nucleic acid binding; IEA.  
DR InterPro; IPR004087; KH.  
DR InterPro; IPR004088; KH\_type\_1.  
DR Pfam; PF00013; KH 1; 3.  
DR SMART; SM00322; KH; 4.  
DR PROSITE; PS00084; KH TYPE 1; 2.  
SQ SEQUENCE 1139 AA; 131625 MW; 9983227714206714 CRC64;

Query Match 4.5%; Score 230.5; DB 2; Length 1139;  
Best Local Similarity 18.9%; Pred. No. 0.073;  
Matches 219; Conservative 173; Mismatches 422; Indels 347; Gaps 45;  
QY 44 GADYAESGSKIKINETSQVDDTVTDLFSDKRTTPPEKIKONLAKPREQSL----- 96  
Db 2 GRKATTSLIEKPKVKNKSGVNNMKDMMDKQGGSKL--NEKKGSKVTSLTNAHNV 59  
QY 97 ---KAVTEN--TESEKQITSGSLE-----OSKESLSLNTKTPVPSNWEI 136  
Db 60 GSIQKKEELENKLNHNKLEENCVKSKEVVGQDKNEGQAKKNNNNNN--- 116  
QY 137 CDPITKGNLVLGSKSGVEKLSQTHLVLPQADGTLQIVASFAFTPDKKTAEYTS 196  
Db 117 -----NKKGTGTENDMDKIQSGNTDNDKKKANKNTTE-----GNEKNKSGNNKKYVG 165  
QY 197 RAGENGESQLDVGDKEI-----INEGEVFNLSYLLKKVITPTGYKHIGODAFVDNKN--- 248  
Db 166 NKDENMKVELMDVTYKNNSTGTGINSSNNNNNNNNNNNNNNNNNNNNNNNNNNNN 225  
QY 249 -----IAEVNLPSELETISDYAFAPALHAKQIDLPDLNKAIGELAPF 289  
Db 226 IISNNNNNNNSDVKKDKWEIKELKEDERKSENDEKKFKV-LNKADNYKINDKK-- 282  
QY 290 DNQITGKLSLQRLMLRAERAFKSNHIIKTIEFRGNSLKVIGEASFQNDLSQML 344  
Db 283 EAEIKKKEI-----ENFISTNSTSRFNGQTTPTI-----NMKLSADM 323  
QY 345 PDGLEKIESEATGPGDDHYNRVVLWTKSGNPSGLATENTYVNPDKSLWQESPEIDY 404  
Db 334 ENKCKDAKSKKKKINP-----KNLEEIEEANTYIN----- 353  
QY 405 TKWLEEDPTYQKN--SVTGFSSNKGIL-----QKVKN-----KNLE 437  
Db 354 --YLEEQSLITKYNENFNKFNQRLISLRKICEELKENLKSNNETKIOEKKLRFVDKLIK 411  
QY 438 IPKHQNGVTITEIG-----DNA-----FRNVDQNKTLRKY-----DLBEVKLPSTIR 480  
Db 412 MKGKQNVDTITEADIIIRKDYAIPDTYTNLMKPNFLNRIQTIFYVYVNDNKLSTNNNN 471  
QY 481 KIGAPAFOSNN-----LKSFE-----ASDLE-----EIKEGAFM 510  
Db 472 MNNNNNNNNNNNNNNNTFINSYDNKTNLHLAGCKDDIEQFISYLNQVNFNSENKFNVHL 531  
QY 511 NNRIETLEKDKLVITIGDAAPHINHIYAIVLPESVOEIGRSAPRQNGANNLIFMGSK--- 567  
Db 532 NKKIPKI-----MLNWDGSP-----KQWEDTVVVFVHVDNMLLYCYCKKDDI 574  
QY 568 ---VKTLAGEMAFSLNR--LEHLDLSEKQKLTETPQAFSDNALKEVLLPASLKTIREEAFK 624  
Db 575 ENLKELIEKATNNEKNLKNKNIKDKLSILARGFNKSKILKEI-----ETNTNTLIRM 628

```

QY 625 NHLQLEVASALSHTAFNALDNDGDEQ-----FONKVVVKTHN-SYALA 669
D 629 NYDQTFDASATIGSKNS-DIBEAEBKLNELKNLDSEFPFEBKEIFSYKKCSYELN 687
QY 670 DGEH----FIVDPDKLSSTIVOLEKILKILIEGLDYSTLRQTQTOTQFDMTTAGKALLSKS 725
D 688 EIRKVLNLFITRODNGLSVGGKONITALEILEYSKNIISNKTVKKLTTEE-EAPLFS 746
QY 726 NLRQSEKQFLOEAQFFL-----GRVDDLKAIKAFKALVTYKATKNGQLLERS 774
D 747 NYRGQIKAKTGAENVKIFKTNHKLNLISGNKNDIDKALEMIEELKKRCKT-----QVD 800
QY 775 INKAVLAYNNSAIKANKVRLEKEL-----DLLTGLVEG 808
D 801 INEKVIALLS--KAQIKDEKOTCTSIQINKTSHVAIQYGHEDNIHLAKDVLENVQS 859
QY 809 KGPLAQATWVGQVYLLKTPPLPEY-IGLVYFDKSKLIYALDMSPTIGEGQDAYGN 867
D 860 EGKEG----EGKF-----IPNNLYVTVMVETEHISVIG--KKGRTINKIQEDTFAK 908
QY 868 PILNVEDNEGYHALAVATLADYEGLDIKTIILNSKLSQTSIRQVPTA-----AYH 918
D 909 KI-HIDKENKKIYIHGTQVDAQKEIQKILLRSKEEMNNNNNNSSNGGAFNNSYH 967
QY 919 RAGIFQATONAAAEABQLPPRGTHSEKSSSESANSKORGL----- 960
D 968 ESRVNNRSHNSSSAHHRKSHKTSRSAPRSHKSESCKGYIINTNDEKAPPSLHDVTN 1027
QY 961 -----OSNPKTNRGRH 971
D 1028 MQSKKKNKLSQSNQTTNTKKH 1048

RESULT 15
Q8EWP8 PRELIMINARY; PRT; 3317 AA.
AC Q8EWP8,
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DE Predicted cytoskeletal protein.
OS OrderedLocustNames-MYPI1550;
OS Mycoplasma penetrans.
OC Bacteria; Filumicutes; Mollicutes; Mycoplasmatataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RC STRAIN=HF-2;
RX MEDLINE=23254719; PubMed=1246555; DOI=10.1093/nar/gkf667;
RA Sasaki Y., Iihikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Morino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
DR EMBL; AP004170; BAC43946.1; -.
KW Complete proteome.
SQ SEQUENCE 3317 AA; 385843 MW; 0641151BAPFA922BB CRC64;

Query Match 4.5%; Score 229; DB 2; Length 3317;
Best Local Similarity 20.1%; Pred No. 0.38;
Matches 222; Conservative 199; Mismatches 406; Indels 280; Gaps 53;

QY 57 KINETSGPVDVTVDLPDSKRTTPBKIDNLAGPRE----- 93
D 1782 ELNNIISNIDSEFKLISFSDSTNQIGENLVKDIKHSNKHAILAKWDSIMEMNSKFDRL 1841
QY 94 -QELKAVTENTSEKQITSGSQLEKESLSLNTKTPVTSNWEICDPITKGNLVLGSKS 152
D 1842 IEEMKVQTKOENSINVDNLNIESKNQNE-----EIPNYVSL-----LEKS 1884
QY 153 GVEKLSQTDHLVLPQAAADGTQLIQVAFPTDK-KTAIAEYTSRAG-ENGEISQLDVD 210

```

```

Db 1085 -----TSSLVNEMQQA---QL-----DKEKIKLKELETLIDLQNSBIESLDD 1924
QY 211 GKELINE-----GEVFNSSYLKKVTIPTGY-----KHIGQDAFVDMKNQIAEVNLP 256
D 1925 KNDPIEBIENFLNOKELFNKENFVLENSTIGNIISAFDKKITDLEFNFORKLSNLDL-S 1983
QY 257 SLETSIDYAFALHAKLQIDLPDNLKAIAGELAFPDNQITGKLSLPRQLMRLAERAPKNHI 316
D 1984 SLKNLFDKDDSTLASVINSVEDNLNSYEI-----LRNELDK--ELNKILEDISEANSI 2035
QY 317 KTI-----EFGNSLKVIGEASFQNDLSQLMLPDGKLEKIESE--AFT 357
D 2036 NEINRENSQPKLEEVNSQFQDFR--DVLKTYTEDIHKSENNQIIVDDLEKLEINDEFPSLI 2094
QY 358 GNPQDDHNNRVVLWTKSGKNPGLAT--ENTYVNPDKSLWQESPEID-YTKWLEE----- 410
D 2095 QNLRSENELSNELMDKFNLSGLVERFNSPNONINNKPFDTAKLDEKTKSPENRMD 2154
QY 411 --DFTYQKN-----SVTGFNSKGLQKVRNKNLEI PKHNGVTTITEIGDNAFR---NVDF 460
D 2155 LINFVYELNKYQKNISSY-----IQKIESNN--EFISDLARTLTKENDKLELEYLIDI 2209
QY 461 QNKTL-----RKVDLEEVKLPSTIRKIGAFAPQSNNLKSFEA-----SDDLEE-IKE- 506
D 2210 QNEEIVRLIDENKX-IDEVEYPLTNKK--EALDNKKLNDIESINNINNNFEKIKKL 2265
QY 507 GAFNNRIETLEK-----DKLVITIGDAAPHINHIYAVLPESVQIEGRSAFRQNG 557
D 2266 GENFEKVDSTELQNLGNLNEESDK-KTLDVIRSQLSKNYELLSEE-----FKNNF 2315
QY 558 ANNLI PMGSKVKTLGEM--APLSNRLEHLDLSEOKQLTET-----PVQAFSDNALK 606
D 2316 A-----SLIKSIENLYSSYGKDSLSQIASSELKFCYNEIGISLGISGNKPNDFFLIK 2368
QY 607 EV-----LLPASLKTIREEAFKKNHLKQLEVASALSHTAFN--ALDNDGDEQ-----F 653
D 2369 ELETNKLILASQVDAMNLILOSNKINSKEIVSYENLKGNYANLSQKEKEQALIFKIF 2428
QY 654 D--NKVVVTKTHNSYALADGEHFIVDPDKLSSTIVLEKILKILIEGLDYSTLRQTQTQF 711
D 2429 DELNEIINIQNDIGSLLSEBDYIIDEVQKPTNKIEBENSKNLIQSADFKNLKDYLTSTV 2488
QY 712 RDMWTAGKALLSKNLRQGEKQKPELQEAQFPLGRVLDLKAIAKAEKAL-VTKKATKNGQL 770
D 2489 DNLGNKIEGL--KSDL-----SNKFTSISOHL-----IDDSAGYENILKVEQNUTENYSP 2538
QY 771 LER---SINKAVLAYNNSAIKANKVRLEKELDLTLGLVEGKGPLAQATWVGQVYLLKTP 827
D 2539 LKKDLYSIFENI FNI IDELIKPSDVASFQKIDDLKEEIDNVNPFILASV----- 2587
QY 828 LPLPEYVIGLVNVPDKSGKLIYALDMSDTTIGEGQDAYGNP--ILNVDE-----DN----- 876
D 2588 -----DKLNKKNQIILDEIQAVITSGKCEANSNDQNYLKVEESVKINDNLELK 2633
QY 877 -----BGYHALAVATLADYEGLDIKTIILNSKLSQTSIRQVPTAAYHRAGIFQAIQNA 930
D 2634 NKISKLDNWKDLVASVENKNGV--ISSFVNKLEBEISKINEVTPAPSDINAYNLLISKI 2692
QY 931 ABAEQLLPKPOTHSEKS-----SSSEGSANSKDRGLQSNPKTNRG--RHSAILPRTGS 980
D 2693 SEISNLL-----NLSKKNLTGLINTVNARLSSESNDKVLQALKWFKNTILFINKTIEE--- 2745
QY 981 KGSFVYGILGYTSVALLSLITAIKKKK 1007
D 2746 -----LAVENNNKYLELINANKNE 2764

RESULT 16
Q7RS28 PRELIMINARY; PRT; 1523 AA.
AC Q7RS28
DT 01-MAR-2004 (TremBLrel. 26, Created)
DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)

```



01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Hypothetical protein.  
GN Name=PY00538;  
OS Plasmodium yoelii yoelii.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=73239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=17XNL;  
RX PubMed=12368865; DOI=10.1038/nature01099;  
RA Carlton J.M., Anguolli S.V., Suh B.B., Kooij T.W., Perteau M.,  
RA Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
RA Peterson J.D., Pop M., Kossack D.S., Shumway M.F., Bidwell S.L.,  
RA Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,  
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,  
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
RA Carucci D.J.;  
RT "Genome sequence and comparative analysis of the model rodent malaria  
RT parasite Plasmodium yoelii yoelii.";  
RL Nature 419:512-519 (2002).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AABL01000149; EAA16762.1; --  
DR InterPro; IPR003123; VPS9.  
DR Pfam; PF02204; VPS9; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 1523 AA; 177622 MW; 7F8FFA7218E933B6 CRC64;  
  
Query Match 4.5%; Score 227; DB 2; Length 1523;  
Best Local Similarity 19.3%; Pred. No. 0.16;  
Matches 202; Conservative 150; Mismatches 410; Indels 284; Gaps 46;  
  
QY 16 SVVTHNQEVFSLVKE-----PILKQTOASSISGADYAESGSKLKINETSQVFD-- 66  
DB 111 SLWESNESYIVKVKCKERNKTPIKKQVNSDKLISS-----MNEISNPFVDQ 158  
QY 67 -----DTVTDL-----FSDKRTPEKIKNL---AKGPREQE-- 95  
DB 159 SNGQKLPDSTPDLWEKTEYDKBYELTGNETSFYBKKKKKKLDYIIHSGKGLSNNSI 218  
QY 96 -LKAVTENSEKQITSGSLEOS-----KESLSLNK----- 126  
DB 219 YKKGTEGENNENVENGENENRIRVMSKSNRVKGDKEIKNSIDYHKCKSKQNTL 278  
QY 127 -TVPSTSNWEICDFITKGNLVLGSKSGVEKLSQTDHLVLPQQAADGTQLIQVASFPTP 185  
DB 279 TKIDSFYBEEFTSLKNGKL-----NDDINLISEHD-IQIECYSDKISQCKVNVNKISI 333  
QY 186 DKKTAI-----AEYTS-----RAGENGISQLDVDGKEIINEGEVFNLYLKQVTI 231  
DB 334 KKKDILLGDEYKNIYKKNKIDIFINSNDNNNNILSDSKHSISNLQINDKRAIKKTKES 393  
QY 232 PTG-----YKHIGQDAFVDNKNIAEVLNPSL-----ETISDYAFALKALQIDLPD 278  
DB 394 PMSNTYSRYRIGNDNNGNDNDLEKTKKPKKNISTQELINSYK-----DCTN 444  
QY 279 NLKALGELAFFNQITGKLSLPQMLRAERAFKSNHKTIEFRGNS---LKVI--GBAS 333  
DB 445 NSVDI-----LVDEQSKDEIQAKREIK-----KKKIMKPINNKTSQSFTSVIDSENT 494  
QY 334 FQNDLSQLMLPDGLEKIE--SEAP-TGNPGDDHNNRVNVLWTSG-----KNPSG 381  
DB 495 YYDNTKFSQSSNLLKLDSESEFFCNSNMWDEKKKKIK--TKKGTDKXNLKXKKKMD 552  
QY 382 LATENTYNPKSLWQESPEIDYTKWLEDFTYQKNSVTGPNKGLQVK---RNKL-E 437  
DB 553 MDNNKKIEIMEIEIKNKSEISQAEOSYSFESIKTNNNSKSFDEIKYLNKNKTKE 612  
QY 438 IPKQHGVTITIGDGNFRNVDFQNTLRKYDLEEVKLPSTIRKIGAFQSNLNKLSFEA 497

613 TNYNYSI-----DDINENFKLNKTL-----EIKKQTIK-----TNSKNTIC 651  
QY 498 SDDLEIEKEGAFMNNRIETLELKOKLVITGDAA-----PHINHIYALVLPESVQIEGRSAF 553  
DB 652 SSDKEIKRNNNSKEKJUVKTKERIKVIGDIADELEIEISNETYMKRENVE--KKGTI 710  
QY 554 RQNGANNLIFMGSK-----VKTGLGEMAF-----SNRLEHLDLSEKQLTETIPVQAFSDNAL 605  
DB 711 KQKGNENIKTNRDKKITENVKKISSEKPYDNSNSTWNSNFSNTYSNLAKKKKKKISDK-- 768  
QY 606 KEVLLPASLTKIRBEAFKQNHKLQLEVASALSHTAFNALDDNDGDEQDFNKVVKVTHNS 665  
DB 769 -----ITVLHSTPNLSRINTNLIEENKNKDCDEKEI----- 800  
QY 666 YALADGEHFIVDPKLSSTIVLEKILKILBGLDYSTLRQTQTQTFQDMTTAGKALLSKS 725  
DB 801 YA-----DILVD-DK-----IIOKENIK-----KNDIKNVYIRDIQDTGTANSKRF 841  
QY 726 NLRQEKQKFLQEAQFFLGRVLDLQKAIKAKALVTKKATKNGKOLLERSINKAVLAYNNS 785  
DB 842 IVK--KKNDFSIDGNETCGQEVDEVEKEEK--KEILENGQNVQEVQNVQESF 897  
QY 786 AIKKAN-----VKRLEKELDLLTGLVSGKGLAQATWV-----QGVYLLK 825  
DB 898 VIQGRKPKMFMFISFIKRESYKQBEVNEQVDENKIITKKEFIKWNEMNLINYNNENKEK 957  
QY 826 TPLPLPEYVI---GLNVYFD--KSKLIYALDMSDTIGEGQKDAYCNGPILNVDENEGVH 880  
DB 958 TESNLRFEPIKVKSTNNELNINSKVI-----EKYKNGLNN--EDNKIFN 1001  
QY 881 ALAVATLADYEGLDIKITILNSKLSOL 906  
DB 1002 NDTISENTSCKKANTYNNFLES 1027  
  
RESULT 17  
Q8GLH0 PRELIMINARY; PRT; 178 AA.  
AC Q8GLH0;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Immunoreactive protein Sello (Fragment).  
OS Streptococcus equi.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=11336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CF32;  
RA Qin A., Artiushin S., Timoney J.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY137519; AAN18285.1; --  
FT NON TER 178 178  
SQ SEQUENCE 178 AA; 19194 MW; 20506DF6D5608BEA CRC64;  
  
Query Match 4.4%; Score 225; DB 2; Length 178;  
Best Local Similarity 35.7%; Pred. No. 0.0096;  
Matches 66; Conservative 35; Mismatches 72; Indels 12; Gaps 6;  
  
QY 1 MKKHKLTVALTLTVSVVTHNQEVFSLVK-EPILKQTOASSISGADYAESGSKSKLIN 59  
DB 3 IKKCPQIALLFLAALASQQTVALERQTSVKQAHESDDW-----FDEGEVSDIKSN 58  
QY 60 ETSQVDDVTVDLFSF--KRTTPEKIKDNLAGPREQELKAVTENSEKQITSGSLEQ 117  
DB 59 -TPTEIDSTVTFLFGDAEPKTKAEK-BESKAPSPQEAALPA---ESQKEKETSASTDHA 113  
QY 118 SKESLSLNKTVFSTSNWEICDFITKGNLVLGSKSGVEKLSQTDHLVLPQQAADGTQLIQ 177  
DB 114 TVQAEAEQPEQSDSPWLADDFTVKGDTLVGLSKAGLAKLSKTPVLVLPFRIGDKITVLR 173







```
QY 75 -----DKRTTPEKIKDNLAQPREOELKAVTENTSEKQITSGSQLBOSKE----- 120
DQ 123 TGRIVYPSSTIPSSIKIKQKGGHSGAKTII-----FDKGSQLEKIEDRAFDFS 174
QY 121 -----SLSLNK-TVPTSTNWEICDFITKNTLVGLSKSGVEKLS 158
DQ 175 ELEETELPASLEYIGTSAFSPSQKLKLTFFSSSKLEL-----ISHEAFANLS 222
QY 159 QTDHLVLSQAD-GTQLIQA-----SPA-----FTPKKTAIABYTGRA 198
DQ 223 NLEKLTPLSKVTKLNLFRLTTSUKHVDVEGNESPASVDGLVPSKD-KTQLIYPSQ- 280
QY 199 GENGISQLDVGKEIIEGEVFNLYLLKVTIPTGYKHIGDQAVDNKNIAEVLNPSL 258
DQ 281 -KNDESYPKPKRELAYSFNKNSY-LKLELNEGLKIGTTFADAIKLEELSLNLS 338
QY 259 ETISDYAF-AHLAKQIDLPNLKAIGE-----LAFDNOQITGKLSPLRQLMLAE 308
DQ 339 ETIERLAFVGNLELAKELILPDNVKFGKHWVGNLKPFLTSGNNIN---SLPSPFLSGVL 395
QY 309 RAFKNHJK-----TIEFRGNS--LKVIGEASQ-DNDLSQMLPDGL 348
DQ 396 DSLKEIHKNSKSTEPSVKKDTFAIPETVKFYVTSHEIKDVLKSLNSTSNDI-----IVEKW 451
QY 349 EKIESEATGPGDDHYNRVVLTWK-----SGKNPSGLATENTYVNPDKSLW-- 396
DQ 452 DNKETDTVAKP-KKNSQGVGVGWKGLWYLNESGSMATGWV-----KDKGLWY 503
QY 397 -QESPEIDYTKWLEED--FTYQKNS---VTGP-SNKGL 427
DQ 504 LNESGSM-ATGWKDKGLWYLNESGSMATGWVKDKGL 540
QY 39736 PRELIMINARY; PRT; 3257 AA.
AC Q9V736;
DT 01-MAY-2000 (TreeBLrel. 13, Created)
DT 01-MAY-2000 (TreeBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TreeBLrel. 26, Last annotation update)
DE CG12864-PA.
GN ORFNames=CG12864;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S.E., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter S.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrez C., Ferreira S., Fleischmann W.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
```

```
RA Merklov G., Milahina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Poulos D.R., Pacle J.M.,
RA Palazono M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G.G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaninker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Bayraktaroglu L., Berman B.P.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Millburn G.H., Prochnik S.E.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003814; AAF58230.1; -
DR FlyBase; FBgn0026427; Su(var)2-HP2.
DR GO; GO:0005720; C:nuclear heterochromatin; IDA.
DR GO; GO:0016458; P:gene silencing; IMP.
DR InterPro; IPR000637; A-T hook.
DR Pfam; PF02178; A-T hook; 1.
DR SMART; SM00384; A-T hook; 2.
SQ SEQUENCE 3257 AA; 355970 MW; 7E9212C20A040810C CRC64;
Query Match 4.3%; Score 220.5; DB 2; Length 3257;
Best Local Similarity 18.9%; Pred. No. 0.87; Mismatches 185; Indels 329; Gaps 51;
Matches 221; Conservative 185;
QY 2 KKHLLKTVALLTITVSVVTH-----NQEVFLSVKKEPI--LKQTOAS-----SSISGADYAES 50
DQ 1206 KKEAVVGVPLDKTSTSSNTIIDKKSNSFDSQMSQPSDRLNQKESAFTKLSSISSPKKIMK 1265
```

QY 51 SGKSKLKINETSGPVDVTVDLFSDKRTTPEKIKONLAKGPRE--OELKAVTENTESEKQ 108  
Db 1266 DQDKOLDALSKGDSNPTIRDTGDSRQTDKKHOENDTKHEREDSSKLANIDETKSSSE 1325  
QY 109 -----ITSGSQLEQSKESISLKNKTVSTSNWELCDPITKGNLTIVGLSKSGVEKLSQTDHL 163  
Db 1326 KDAEPIKSDOSSQSAKPRISKPKSRNKRKNE-----KKPND--STAESDIEGGFQVN-- 1376  
QY 164 VLPSQAADGTQITQVASFATP-----DKTAIAEYTSRAGENGIEISOLDVDGKEIINEGBV 220  
Db 1377 -----TETVQAT--CSTSESNNKDMVK--SDETNEEPNLSETEIG--RIRKGOA 1421  
QY 221 FNSYLLKKVITPGYKHGQDAFVDMKNIAEVLNPSLETISDYAFAPHLALKQIDLPDLN 280  
Db 1422 FH-----IENPKDDLHITPON--ENQSIAGVNF-----KQVLPFESV 1457  
QY 281 KAIGELAFDNIQTGLSKLPRQLMLAERAFKSNHIKTIEFRGNLSKVTGGEASFQNDULS 340  
Db 1458 ES-----DTPI--MKIPTKTYLMCTK-----NKTSL--SASEDPDI- 1490  
QY 341 QLMPLDGLKISEAFPTGNPGDDHNNRVVLMTKSGKNP----- 379  
Db 1491 --VSEPOKL--ITTSKGDSPDLNANN-----LETSSQDPKEHEFSQDTTDSNDIIPST 1544  
QY 380 -----SGLATENTYVNPDKSLWQBSPEIDYTKWLEEDFTYQKNSVTGFSNKG 426  
Db 1545 KKSQIVFPPTTKSSDOTKNSFITPNRSPKSRNVSKEAKRLDNSPEESQNAASESSASK 1604  
QY 427 LOK-----VKNQKLEIPKOHNGVITTEIGDNAPRVNDFONTKRYDLE 471  
Db 1605 VQELRTPTASCRKLRLVLRKTPTSLSL-----TNSRKSTFKKTPAKSRRLKILES 1656  
QY 472 EVKLPS-----TIRKIGAF-----FOSNLSKFEASDLEIEKEGAF 509  
Db 1657 MEKTSREPSVSLGEVNPDPDVAESAVALHESDRLESNEIPNEVEFDETEE--ASABD 1715  
QY 510 MNNRIETLKLKLVITIGDAAFHINHIYAIVLPSVQETICRSAFRONGANNILIFGSKVK 569  
Db 1716 TDKLKKKE-----DHELEVDNICAASKNPITDDSTKQASSNSTDSVLQETKDE 1767  
QY 570 TLGEMAFNLNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREAFKNNHLKQ 629  
Db 1768 -----LNSNLINATQGEDTPIKELTEBEPNN-----KTVEDESKQBEILKQ 1809  
QY 630 LEVASALSIAFNALDDNGD-----EQFDNKVVVKTHTNSVALADGEHFIIDPKLS-S 683  
Db 1810 LEPDNA-----ALEEDTASTAKAAEEMDLTYKEKSNVKS-VLAEPETDVTDDDELAQS 1861  
QY 684 TIVDLKILKLEGLDYST--LROPTOTQPRDMMTTAGKALLSKNLRO-----GEKQ 733  
Db 1862 PIPNSETTSVTDPEPSTSSVVKSLRLEADSSQPDDEA--AKRKQROVEKSLTGKKE 1919  
QY 734 KFLQEAQFFLGRVLDLKAIAKAFKALVTYKATNGOLLERSINKAVLAYNNSAIKKAN-- 791  
Db 1920 QVKPARRQLAEVE--ERPSLKRKSTSEAKSTVGKYISIIAGNETIMSTTAPIRETNR 1978  
QY 792 -----VRLKELDLTG-----LVEGKGFLAQAATWQGVYLL 824  
Db 1979 AASTSPSARKSAVQEAHVKHETTKHIIILGPPGKLLHSDSPAEEVKKPMVQ--TLLSSTLSL 2037  
QY 825 KTLPLPEYIIGLVYFDKGLIYALDMSDTIGEGQDAYGNPILNVDEDNHGYHALAV 884  
Db 2038 QKPSTL-----DGSPKLRKSLKKSSTADE-----NIDGQISFSSSV 2076  
QY 885 ATLADYEGLDIKTILNSKLSQTS--IRQVPTAAYHRAGIFQAIQNAABAEQOLLKPKGPH 943  
Db 2077 --LNKNTSVAPKVNISVLSQSKDTQVETAA-----SSSTPILTKEKUL 2122  
QY 944 SKGSSSESAN-----SKDRGLQSNP--KTNRGHRSAILPR-----TGSKGSFVYGI 988  
Db 2123 TQKSTKPKGNKTESKKSLSVQGPOMKTOKEAASVSGPKILNKYLKSETTESSRKTVSTV 2182

QY 989 LGYTSVALL-----SLITAIIKKK 1007  
Db 2183 TGRKQIGQLEVLKPKSRKSESLVEAISRK 2214  
RESULT 22  
QYVJQO  
ID QYVJQO PRELIMINARY; PRT; 1443 AA.  
AC QYVJQO, Q9NK96,  
DT 01-MAY-2000 (TREMREL. 13, Created)  
DT 01-OCT-2002 (TREMREL. 22, Last sequence update)  
DT 01-MAR-2004 (TREMREL. 26, Last annotation update)  
DE CG4168-PA.  
GN ORFNames=CG4168;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Cocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Chertys J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Gelbart W.M., Glaeser K.,  
RA Fosldek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milnina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler P., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195 (2000).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2426065; PubMed=12537568;  
RA Celnik S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila  
melanogaster euchromatic genome sequence."  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).  
[3]





Db 1491 SKLDOEIKTKKEAEPNNLTNAKTELSNLTQAPAEALSKAKKVLVEEINGLNLSDSIK 1550  
 Qy 702 TLRTQTOQPRDMITAGKALLSKSLNRQEQKQFLQBAQFFLGRVLDLK-----AIAKAE 756  
 Db 1551 SLKATO-KIRDAETQLTAIEKT-----EKLKFEDEARKSLRELKIDDDAKAEVETSKAK 1604  
 Qy 757 KALVTKATKNGQLLERSINKAVLAYNSAUKANVKLEKELDLTLGLVGGKGLAQAT 816  
 Db 1605 EALKTKAKDENSESTEE--IIKATKALEDAKFKELNOEIRVK--DRLMKLNKAKKKLDA- 1659  
 Qy 817 MVQGYLLKT-----PLPLPEYIGLVNFYDSKGLIYAL-DMSDTIGBGKXDAYGNPILN 871  
 Db 1660 -----LLKTENLDDLSIEKLGIS-----AGEIKNKLEDAKLLQEAAGLSSENSKITN 1708  
 Qy 872 VDENEHGHALA-----VATLADYEGLDIKYILNSKLSQLTSTR 910  
 Db 1709 IKOKIQIKIDYVROPEPWISNMLKIK-GLSVINSNIEQLKKFK 1752  
 RESULT 25  
 PCPI\_SCHPO STANDARD; PRT; 1208 AA.  
 AC Q92351;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Spindle pole body protein pcpi.  
 GN Name=pcpi; ORFNames=SPAC6G9.06c;  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_taxid=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=21852775; PubMed=11864908;  
 RA Flory M.R., Morpew M., Joseph J.D., Means A.R., Davis T.N.;  
 RT "Pcp1, a Spcl10p-related calmodulin target at the centrosome of the  
 RT fission yeast Schizosaccharomyces pombe.";  
 RL Cell Growth Differ. 13:47-58(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=2184401; PubMed=11859360; DOI=10.1038/nature724;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne K., Lyne K., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jørgensen K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Mølle S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weitzens I., Vanterreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leinhardt R., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux C., Lelaire V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;  
 RT The genome sequence of Schizosaccharomyces pombe.";  
 CC Nature 415:871-880(2002).  
 CC -1- FUNCTION: Spindle pole body component that binds calmodulin.  
 CC Overexpression of pcpi causes the formation of supernumerary SPB-  
 CC like structures and disrupts both mitotic spindle assembly and

CC chromosome segregation.  
 CC -1- SUBCELLULAR LOCATION: Spindle pole body.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/  
 CC or send an email to license@ebi.ac.uk).  
 CC -----  
 DR EMBL; Z81317; CAB03608.1; --  
 DR EMBL; AF348506; AAK31344.1; --  
 DR PIR; T39068; T39068.  
 DR HSSP; P08670; IGK4.  
 DR GeneDB SPombe; SPAC6G9.06c; --  
 KW Calmodulin-binding; Coiled coil.  
 FT DOMAIN 151 375 Coiled coil (Potential).  
 FT DOMAIN 387 803 Coiled coil (Potential).  
 FT DOMAIN 874 1091 Coiled coil (Potential).  
 FT DOMAIN 1177 1204 Coiled coil (Potential).  
 SQ SEQUENCE 1208 AA; 140763 MW; 70264159ADD42424 CRC64;  
 Query Match 4.3%; Score 219; DB 1; Length 1208;  
 Best Local Similarity 19.8%; Pred. No. 0.25; Indels 278; Gaps 50;  
 Matches 227; Conservative 196; Mismatches 443;  
 Qy 32 ILKO-TQASSISGADYAESGSKLKNETSGPVDVTDLFSDKRTT---PEKIDNL 87  
 Db 104 ISKQATQEALSISQGN--DSVDVSKTDLKSNSEIDHTDGLPANAALTREQEKVLEKV 161  
 Qy 88 AK---GPR-----EQLKAVTENTSEKQITSGSQSLESLSINKVTP 129  
 Db 162 SRENFGRLRIKIVCLEKLESMAPEQIKAEVKONVELHAE-RANLQQLKRTESLQKS-- 218  
 Qy 130 STSNWEI---CDFITKGNLT-----VGLSGVKEKLSQTDHLPLPSQAADGTQ 174  
 Db 219 EDKMFLEKVDYLSKNDVEQSONVQVFTIRIFLENALEKVKOREKDSISTEED--- 275  
 Qy 175 LIQVASFAPTPDKKTAJAEYTSRAGEN--GEIS-QLDVGDKIIEGNEGVNFSYLLKVTI 231  
 Db 276 -----KSNKEVDYVEIRQLQNRDLSELDV-AQDLLTEKEDETATLKQIE- 323  
 Qy 232 PTGYKHIGQAFVDNKNIAEVLNPLESLETISDYAFALHAKQIDLPNLKAI-CEL-AFF 289  
 Db 324 ---EKENSsafeneerssyvhlqe-----DYAI--LQAKCDEFADRIQVLTADLEKKB 372  
 Qy 290 DNOITGK-----LSLPRQLMLRAERAPKSNHIKTIEFRGNLSKVIGE--ASFQNDLSQ 341  
 Db 373 ENQIMHSEASIGLTDMSQVHTLOEQHLKAN--BEIEFLHDDQISRMNEEGKNFEDIMLQF 430  
 Qy 342 LMLPDGLEKIESEAFNGDDHYNNRVLVLTWS--GNPGLATENTYVNPDKSLWQESP 400  
 Db 431 RSLSEERDVLESKLQTL--DD--NNSLRLTWSLGNQIESLRTQNRIDEK----- 479  
 Qy 401 EIDYTKLEEDFTYQKNSVTFGSKNGLOKVKGNKQLEIPKQNGVTTIEGDNAFRN--- 457  
 Db 480 --NHLRLLA-----SKNSKALAEATNIRLOEVTKETLRMKNNDLNEIHLRENEGL 532  
 Qy 458 -----VDFQNTKLRKYDLEVKLPSTIRK-----TGAFQFQ 488  
 Db 533 TLKIDSTTKEDRLINLEBQRIKSYEVNVSELNTGIDIEYRNKLDKKEETYNVMAFOYK 592  
 Qy 489 SNNLKSFEASDDEEIEKEGAFPMNRIETLEKDKLVITGDAAPHINHIYAILVP--ESVQ 546  
 Db 593 DNDLRRPHESINKLQDREKELTSN---LEKNLIVIS-----SLRETVMAMLEKERSIK 642  
 Qy 547 E-IGRSAPRQNGANNLIFMKSQVKTIGEMAFLSNRLEHLDSEKQKTEIPV-----QAF 600  
 Db 643 KYLSGNKADLNTNLMELINDKITSVLQRQ--LTDVKDELDSVSEEREBAIVAGOKLSASF 700  
 Qy 601 SDNAKVELVLPASLKTIREEAFKKNHL--KOLEVASALSHTAFNALDNDQ---DEQPDN 655



```
Db 701 ELMSEKQALELKYSLKNEVLNAQNLLDRREELSELSEKSLFPERKIRSGNDDIEKNK 760
QY 656 KVVVTHNSYALADGEHIVDPDKLSSTIVDL-----EKILKLEGLDYS 701
Db 761 EINVLSELADKLAQIRHLESQWELDKLVHNLRGIEBANIEANAKKRLCLMGCDYS 820
QY 702 -----TLRQTQTQFRDMWT-AGKA----- 720
Db 821 SVSILQIVSQIEHFVNQQTIRSLKQELRHDFVQSGKQEQLSRSPEKFGLTETKHD 880
QY 721 LLSKSNLROGEKQKLOE-AQELGRVD-----LDKAIKAEKALVTK 762
Db 881 ILAQRNNRVSEKNDLENAAQFFSPDRKNGLYPSEHTSKIEYLEKTIEDLKAL--Q 938
QY 763 KATKQQLERINKAVLAYNNSAIK-KANVRLEKELDLLGLVEGKGPLAQATWVGQV 821
Db 939 DELKRNLLMDIIS-----SYNKQTYKQEKIKWLERERSILDELES----- 981
QY 822 YLLKTLPLPEYVYIGLVNFDKSLIYALDMSDTIGEQQ--DAYGNPILNVDEDEBY 879
Db 982 -----YRSNQFNQNNLVQDKN-----ELEERLKEIQEKELEVYNNHFMQAE----- 1023
QY 880 HALAVATLADYGLDITILNSKLSQTSIROVPTAAHYRAGI FQAIQWAAAEAEQLLPK 939
Db 1024 --LMTSNVTDESQMLKTLREALQSKTNNDHLSTILERNRKEYKSLDDYVNLRAKYK 1081
QY 940 POTHSEKSSSSSANSKDRQLSNPK---TWGRHSATLPRTGSGSPVYGILGVTVA 995
Db 1082 LOSNTPQSTQSQYSEIKGLSKLYKLOSKCRREHSLDLAFSKKFLMQLTGYTCN 1141
QY 996 LLSL 999
Db 1142 KINL 1145

RESULT 26
Q7RHM8 PRELIMINARY; PRT; 1675 AA.
AC Q7RHM8;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Protein kinase domain, putative.
GN Name=PY03956;
OS Plasmodium yoeelii yoeelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.B.,
RA Shallow S.J., van Aken S.E., Feldmuller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoeelii yoeelii."
RL Nature 419:512-519 (2002).
CC -!- SIMILARITY: belongs to the Ser/Thr protein kinase family.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AABL01001181; EAA15751.1; -.
DR HSP: O14965; 1MQ4.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
```

```
DR GO: GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR011591; Botulinum.
DR InterPro: IPR011009; Kinase like.
DR InterPro: IPR000719; Prot kinase.
DR InterPro: IPR008271; Ser_thr_pkin_AS.
DR Pfam: PF00069; Pkinase; 1.
DR ProDom: PD001963; Botulinum; 1.
DR ProDom: PD000001; Prot kinase; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; UNKNOWN_1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1675 AA; 195244 MW; 281BE2B42E5DD60 CRC64;

Query Match 4.3%; Score 218.5; DB 2; Length 1675;
Best Local Similarity 19.2%; Pred. No. 0.42;
Matches 228; Conservative 192; Mismatches 466; Indels 303; Gaps 51;

QY 21 NOEVPFLVKEPILK---QTQASSISG-----ADYAESGSKSLKINETSQFVDDTVTD 71
Db 18 NKVPSNLVNDLILKFNFTNI KRNNVVERFVHNKVSDDSNMYNKNKGHHINKSVG-----AEKN 73
QY 72 LFSDKRTTPEKIKONLAGPREQELKAVTENTSEKQITSGSQLESLSLNTVPSST 131
Db 74 ISGNKTNLVAIKN-----KNQRLFSKSEN---EKVIRNINREDNYTTNLIK----- 119
QY 132 SNWEICDPITKGNLTVGLSKSGVEKLSQTDHLVLPSSQAADGTOLLIQVAFAPDPKTAI 191
Db 120 -----GKNDIINNRKN--TWNNVEYLNLSKSDDKLTNINTSGIAT-----ENI 163
QY 192 AEYTSRAGEGEISQLDVDGKEIINEGVFNFSYLLKVTIPTGYKHIGQDAF-----VD 245
Db 164 LEDNCNNGNDIRNDIEKDNNGNNGYYSCKNLKKVKEPNN-----SLDSFYKREKRETE 219
QY 246 NKNTAEVNLPESE-----TISDYAPAHALAKQIDLPNLKALGELAFPDNQITGKL 297
Db 220 EENYDKKRINHSLNKDYYNTYTKTDINDSHMNIKDIQIKOISENVSKIKLENNIMNVK 279
QY 298 SLPRQLMELAEAPKPSNHIKTI--EFRGNSLKVIGEASFQDN-----D 338
Db 280 NWEKARSVHNISESEKKNFNIINQSSNNSTKILNHSFNTNESCNVYKMKDI FNKNVD 339
QY 339 LSQMLPDG--LE-----KIESEAFGTGNPDGDH-----YNNRV---VLMTKSGK 377
Db 340 YKKIKNPENSVLETDENIISRSYKLEEEHI--NNSDDFIRNFWYDNKLETINKTSNK 397
QY 378 NPSGLATE-----NTYVNPDKSLWQESP-----EIDYTKW-----LEEDFTYQK 416
Db 398 LEKNILTDIKSKNSYILKHKQYNEKGNFCDNEFKQNKYDFISCSFNEEEEFMNDY 457
QY 417 NSVTGP---SNKGLOK-VKRNKNLE-----IPQHNGVITITEIGNAFRN----- 457
Db 458 INNTDFVKNKEGLNKLIDINKNEBEQNAAFITLKNENNINNETOBYNNSHENIKULLYNNV 517
QY 458 -----VDPQNTKLRYKYLEEVKLPSTIRKIGAFQFQSNLKSFEASDLEETKEGAF--- 509
Db 518 KDAEIEFNDISLNDPKMDVSKPIDTKNTKISQYQESNII--YKINNDNEAKDNWEYD 575
QY 510 ---MNNRIETLEKOKLVITIGDAAPHINHIYAIIVPESVQEI---GRSAFRONGAN----- 559
Db 576 LSSNN---CKNQNDGLKTIIG-ALANDIMNTYKRLVDNTNNVNLNKGQKSDSENLNKK 631
QY 560 -----NLIIFWG-----SKVKTLL--GEMAFSLNLEHL 585
Db 632 RTEKKNVALDHKNRVNERNLIISASNTNNNSNVSVKRVNNSGDISSTSENKKNL 691
QY 586 LSEOKQLTEIPVQAF-----SDNALKEVLIPASLKITREAEAFKQNHKLQLEV 632
Db 692 IKKTYNESISKRLPSTHLSNNKCLSSGNYLKNKIKP-SIKNSFKTYKKNLDRHKT 750
QY 633 A-----SALSHIAFNALDDNDGDEQFQDNKVVVTKTHNSYALADGEHFVDPDKLSSTI- 685
Db 751 IGFNRSNTLNDKINSNQNGITKNSIQKPPSSNIHTNLSNIQSKKKTIDNNYKNSTIE 810
```

```
Qy 686 -VDLEKILKLEGLDYSLRQTTQTPRDMTAGKALLSKSNLRQ---GEKQKFLQEAQF 741
Dy 811 NKNLEENNEKIKKQKIBFNKTNNPSFLER-----MKKKNLNEHVDGNKNAFTQKTVQ 863
Qy 742 FLGRVLDLKAATAKA---EKALVTCKATKNGQLLE-----RSINKAVL--AYNNSAI 787
Dy 864 NRNNINFENKTKTKMPLSNKKKITNNYNGKNGLFLSELAEKKISKSTDKLTIDKKYNHSGH 923
Qy 788 KKANVKRLKLEKLDLLTLGLVEGKGPLAQATQGVVLLKTLPLPEYIYIGLAVYFDKSGKL 847
Dy 924 VEVGAK-LEKGL-----GK-----EIEGLQGEISINSKV 952
Qy 848 IYALD-----MSDTTGEQKDAYGNPILNVDENEGHALAVATLADYBGLDIKTLNSKL 903
Dy 953 EEKIEDPLYKQNNVGE---VFRENMDTNEISNINYSKNTNKAKANALNRERINNKK- 1007
Qy 904 SOLTSIRQVPTAAVHRAGIFQAIQNAAEAEQLLPKPTGTHSEKSSSSSASNSKDRGLOS 963
Dy 1008 -----BIRTLASSNS-ITSGKNTLLVKGQNTNNYMYGSKIKIEIDNKLKMKNLANN 1059
Qy 964 PKTNRGHRSALP-----RTSGSKGSFVYGILGYTSVALLSLITAIKKKK 1007
Dy 1060 ---NKGIDNKYVPKSSNKTSSSNI TNIVDKNNMLVKKDEINKKKK 1105

RESULT 27
Q6KIF4
ID Q6KIF4 PRELIMINARY; PRT; 662 AA.
AC Q6KIF4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Putative expressed membrane/lipoprotein.
GN OrderedLocustNames=MMOB1360;
OS Mycoplasma mobile.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2118;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=163K / ATCC 43663;
RX PubMed=15289470; DOI=10.1101/gr.2674004;
RA Jaffe J.D., Strange-Thomann N., Smith C., Decaprio D., Fisher S.,
RA Butler J., Calvo S., Elkins T., FitzGerald M.G., Hafez N.,
RA Kodira C.D., Major J., Wang S., Wilkinson J., Nicol R., Nusbaum C.,
RA Birren B., Berg H.C., Church G.M.;
RT "The complete genome and proteome of Mycoplasma mobile.";
RL Genome Res. 14:1447-1461(2004).
DR EMBL; AE017308; AAT27622.1; -.
DR InterPro; IPR007326; Lipoprotein_17.
DR Pfam; PF04206; Lipoprotein_17; 2.
KW Complete proteome; Lipoprotein.
SQ SEQUENCE 662 AA; 72407 MW; BB9543519515C79D CRC64;
[1]

Query Match 4.3%; Score 218; DB 2; Length 662;
Best Local Similarity 20.6%; Pred. No. 0.12;
Matches 159; Conservative 130; Mismatches 261; Indels 220; Gaps 42;

Qy 123 SLNKTPVSTSNWEICDFTKGTNTLVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVASFA 182
Dy 19 SLTITIVSCSTASISDL-----QLV-ISKNSKITASQNVPASSITSD-EMILQINLKL 71
Qy 183 FTPDKKT---AIAEYTSRAGENGETISQLDVGKEITINEGEVFNPSYLLKKVTIP-TGYKHI 238
Dy 72 TIPSVSTNINLIFQLVNGNANDQGSGLD1--KIATNGTVGNQVISISSQTLTVTGFLTQ 129
Qy 239 GQDAPVDNKNATAEV-NLPSELETISDYAFAPHALKQIDLDPNLKKAIGELAFPDNQITKL 297
Dy 130 TQ-----VNNNTNLEIKNIIFDRFETSLKNGSVISVARTQITN-----ETNPVIGSN 177
Qy 298 SLPQLMRLAERAFKSNHIKT-EPR-----GNSLKVIGEASFQNDLSQLMLPDGLSKI 351
Dy 178 TLP-----SKVSQSLNSTLIQGLQEFQSLNSASNSVNLTEANYIDGSAN-----DALGTI 229
```

```
Qy 352 ESEAFNGDGDHYNRRVVLWTKSGKNP-----SGLATENTYVNPDKSLMGESEIDYTKW 407
Dy 230 QIS-----LQGTIAGNIPTISKSKSVTIGFLKAEKQILLESNK-----NW 270
Qy 408 LBEEDFTYQKNSVTGFSNKG-----LQVKENKNLEIPKQ-HNGVTITIEIGDNA 454
Dy 271 IRD---ISKETIONFANSNKEAFINDQGKULLSTLKNRNPFSLRSSIKNDNLVNYGSPS 327
Qy 455 FRNV-----DFQNKRLKRYOLEEVKLPSTI-----RKIGAFAPQSN 490
Dy 328 FFNTIINSNODYSIVATISGENISNDFEGSVLDFQI-----TIGNREIGTF-FSSS 379
Qy 491 -----NLKSPASDDLEIEKEGAPMNRRIETLELKD-----LVTIGDAAFHIN----- 534
Dy 380 LIEKVLNIGFKKSSSLH-----NTIFVRDKFNRLSVLNTGEISAEELNGPDTN 427
Qy 535 -----HIYAIVLPESVQEIGRSNFRGANNLIPMGSKVKTLGEMAFSLNRLEHDLSE 588
Dy 428 LNPFILEDLTRIVIPSTVEDIGGNIF-----KVKSIKDSAPINSNLRLLDLSN 474
Qy 589 QKQLTEIPVOAFSDNALK-EVLLPASLKTIREAPKKNHLKQLEVASA-----LSHIA 640
Dy 475 ASNLESIGNFSPSNKNIENSIQIPASLKNIGENAFSNLNKINFGNASLSOLEKINKAA 534
Qy 641 FNALDDNDGDEQFQPNKVVVT-----HNSYALADGEHFIVDDPKLSSTIVDLEKILKIE 696
Dy 535 FL-----NNNLESFQVNGVFSNLPFIGNAFANSNPN-----SRLST--ISLDSIINFLN 583
Qy 697 GLDYSLRQTTQTPRDMTAGKALLSKSNLRQGEKQKFLQE--AQFGLGRVLDLKAIAK 754
Dy 584 DFTFT-----QDLTI-----EKGAFLRELKSDFWIPQYNLN----- 614
Qy 755 AEKALVTCKATKNG-QLLERSINKAV-LAYNNSAIK--KANV-KRLEKEL 799
Dy 615 ---ALNRLHVLKNSIKFLSNQNNNGIDTSPGNVAVSNFKTIQEKYEKNI 661

RESULT 28
Q7RR45
ID Q7RR45 PRELIMINARY; PRT; 2341 AA.
AC Q7RR45;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY00892;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Koolij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feidilyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabibi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01000239; EAA19159.1; -.
DR InterPro; IPR003900; KID_repeat.
DR Pfam; PF02524; KID; 9.
```



KW Hypothetical protein.  
SQ SEQUENCE 2341 AA; 266712 MW; 51D2041A084F4BA5 CRC64;

Query Match 4.3%; Score 218; DB 2; Length 2341;  
Best Local Similarity 21.1%; Pred. No. 0.71;  
Matches 207; Conservative 147; Mismatches 396; Indels 232; Gaps 46;

QY 53 KSKLKINETSGPVDDVTDTDFSDKRTTPBKIKDNLAAGPREOEL-----KAVTEN----- 102  
DB 845 KTELANEHISKKWENTINNEKVSNSISEQSIDNATKTENELVLKVSISSENQKQN 904  
QY 103 -----TESEKQITSGS-----QLEQKESLSLNTKTPVSTNSWEICDFYKGNLTVGLSKS 152  
DB 905 PTIITOSKDPNKGSIINIYCDYHENEQKSAANSADNSSEK-----KKTNYLINS-- 957  
QY 153 GVEKLSOTDHLVPSQAADGTQLIQVAFPT-----PDKTAIAEYTSRAGENGISQL 207  
DB 958 --EKLSENEKLNWIDNSKNRANICITNSGDHSNNKNSDITYNGSENS----- 1010  
QY 208 DVDGKEIINEGEVF--NSYLLKKVTTIPTGYKHIGODAFVNDKNIAE-----VNLPSLETI 261  
DB 1011 --SSKEASGLSVFKDSEYSMREESEKTNL-HVNDSEYTSIKMSRMASYINGSEK---- 1063  
QY 262 SDYAFAHALKOIDLPDNLKALGELAFNQITGKLSLPROLMRLAERAFKSNHIKTIEF 321  
DB 1064 -----TLSKKDLKINSR-----IDKSEKALNKKDIEIDSH 1093  
QY 322 RGNLSKVIGEASFQDNDLSQMLPDGLEKIESEAFNPGDDHYNRVVLTWTKSGK--NP 379  
DB 1094 INKSEKILSEGGFEIN--SHINNP-----EKWANE--KGLIEDSHLN-----KSGRMVNE 1139  
QY 380 SGLATENTYVNPDKSLWQSPSEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVKRNKLEIP 439  
DB 1140 KDESDSHLNKSGKMWNEKDEID-----SHLNKS--CKMWNEKDEID 1181  
QY 440 KOHN--GVTTIEIGNAFNVPQNTLRYDLEEVKLPSTIRKIGAPAFQSNLNKSPFA 497  
DB 1182 SHLNKSGKMLNKGLEIDSNISKEKTLNKKGLE--INSHISKSG--KTVNEKDLBS 1234  
QY 498 SDDLSEIKEGAPNNRIETLELKDVLVTIGDAAPHINHIYAIVLPESVQEIERSAPRQNG 557  
DB 1235 DSHLN--KSGKIVNEK--GLEIDSHLNKSGK-----MLNEKGLEI--DSHLNNS 1277  
QY 558 ANNLFMGSKVKTGLGEMAFSLNRLEHLDLSEOKQLTEIPVQAFSDNALKEVLLPASLKI 617  
DB 1278 ENTLSKKGKLI-----NKRINSEKALNKKOLEI-----DSHINK-----SEKIL 1317  
QY 618 REEAFKKN-----HLKOLEVASALSIAFNALDNDGDGEQPDNKKVVVYKTHNSYA 667  
DB 1318 NEEGLKINSRINNSEKTLNKKDPEIDSHL-----NKGKMWNEKGLESDSH--- 1363  
QY 668 LADGEHFIVDPD-KLSSTVIDLEKILKLEGLDYSTLRTQTOTQFDMWTAGALLSKSN 726  
DB 1364 LNESEKMLSEKGLKIDSNISKEKTLNKKGLEIDSNINNSN-----TLSEKGL--KIN 1415  
QY 727 LRQGEKQKFLQBAQFPLGRVDLDKATAKAEKALVTKATK-----NGQLLERSINKAVLAY 782  
DB 1416 SRINSEKALNKD-----LEIDSHLNKSEKTLNKKGLEIDSHLNKSGKMWNEKGLEIDSHLN 1468  
QY 783 NNSAIKKAANVXBLEKELDILLTGLVEGKPLAQATMVQGVYLLKTLPLPLPEYIIGLVNVPD 842  
DB 1469 -NSRINNSEKMLSEKGLKIDSNISKEKTLNKKGLEIDSHLNKSGKMWNEKGLEIDSHLN 1527  
QY 843 KSGKLIYALDMSDTTIGEGQKDAY-GNP--ILNVEDNEGYHALAVATLADYGLDKITLL 899  
DB 1528 KSGKMWNEKDL-----EMDSSHINPEKILDEGGIEINSHINSEKMLSEKGLEIDSHL 1580  
QY 900 NKKLSQLTIRQVPTAAYHRAGIFQAIQNAAEAEOLLPKPQTHSEKSSSSSANSKORG 959  
DB 1581 N-KSGKILSEKGL-----AIDSHLNKSGKILSEKGLAIDSHLNKSEKIPNERG 1627  
QY 960 LQSNPKTNRGHSAILPRTGSK 981

DB 1628 IEINSHINKS--EKILSEKGLK 1647

RESULT 29  
Q9CF64  
ID Q9CF64 PRELIMINARY; PRT; 1072 AA.  
AC Q9CF64;  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Hypothetical protein YqG.  
GN Name=yqG; OrderedLocNames=LL1617;  
OC Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IL1403;  
RX MEDLINE=21235186; PubMed=11337471; DOI=10.1101/gr-1697R;  
RA Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,  
RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
lactis ssp. lactis il1403";  
RL Genome Res. 11:731-753(2001).  
DR PIR; A86827; A86827.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 1072 AA; 113056 MW; 464446E2656CAA08 CRC64;

Query Match 4.3%; Score 217.5; DB 2; Length 1072;  
Best Local Similarity 16.9%; Pred. No. 0.25;  
Matches 151; Conservative 159; Mismatches 365; Indels 219; Gaps 29;

QY 149 LSKSVEKLSQT-----DHLVLPQAADGT-----OLIQVASFAPTPDKKT 189  
DB 1 MSQSIKKTMTVGGLLTATLNPFTLNHQLTELPEATTSKITTYTDEQKITNLDKVS 60  
QY 190 AIAEYTSRAGENG-EISQLDYDVGKEIINEGE-----VFNSVLLKKVITPTGYKHI 238  
DB 61 LTAEVASVAGLAGADIVIPDI-----IVNNGQTYAITSIGTYAFSGIRSVIIGNVVDI 116  
QY 239 GODAFVD-----NKNIAEYVNPESLETISDYAFALHAKQIDLPNLKAIAGELAPFD 290  
DB 117 NTSAFQTTTAPPDYKKSILTQVLGAKVQNIKTDAFAGNAISSIEFPNSVLKIATRAPAN 176  
QY 291 NOITGKLSLPQMLAERAFKSNHIKTIEPRGNSLKVIGERASFOQDNDLSQMLPDGLEK 350  
DB 177 NNLT-ELSLGSGNITEIMAKAFQSNQITTIEFADESLLTIVDSAAFSGSSVQSLTLGIV-- 233  
QY 351 IESEAFNPGDDHYNRVVLTWTKSGK-----NPSGLATENTYVNPDKSLWQBS 399  
DB 234 -----TLADDVFNKTSPLFGQSDLPTEIRTIYSNSSL-PDKSWINSDDSS---OS 281  
QY 400 PEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVKRNKLEIPKOHNGVTI-----TEIGDN 453  
DB 282 TSVS-----TENADTTPVSSSNSVEQVDVASS-ESTQDANSASLYPISEASVTDN 331  
QY 454 AFRVNDPQNTLRYDLEEVKLPSTIRKIGAPAFQSNLNKSPFAED-----DLEE 503  
DB 332 TLNLSISLSDSSISSQSTENSO--SGASSTAEISYDSSENSLSLSSNQINSNSSEKDSNQ 389  
QY 504 IKEGAFMNRLETLELKDVLGTIGDAAPHINHIYAIVLP-----ESVQEIERSAFRONGANN 560  
DB 390 SLSGSMSESESEHSNSNINETNNSSEITNLPSPNPTESNSVSDQTSSEASTNSNS 449  
QY 561 LTFMGSKVKTGLGEMAFSLNRLEHLDLSEOKQLTEIPVQAFSDNALKEVLLPASLKTIRE 620  
DB 450 ISLSPSNISSTSDSSATNSDPSNVAEVANNSLASVNSSSSVLSSTSTADNL-GINQS 508  
QY 621 AFKNHKLQLEVASALSIAFNALDNDGDGEQPDNKKVVVYKTHNSYALADGEHFIVDPDK 680  
DB 509 GSDNLTQDSSEISTSGAFLLSNQTSSE-----ASTNSNS-----SISLSPSN 550







RESULT 34  
 ID COB1 HUMAN STANDARD; PRT; 3259 AA.  
 AC Q14789; Q14398;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macrogolgin)  
 DE (Golgi complex-associated protein, 372-kDa) (GCP372).  
 GN Name:GOLGB1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94187728; PubMed=7511208;  
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,  
 RA Renz M.;  
 RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane  
 RT protein (giantin).";  
 RL Mol. Cell. Biol. 14:2564-2576(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94257116; PubMed=8198703; DOI=10.1006/jaut.1994.1006;  
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,  
 RA Renz M.;  
 RT "Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as  
 RT target of antibodies in patients with rheumatic diseases and HIV  
 RT infections.";  
 RL J. Autoimmun. 7:67-91(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95100974; PubMed=7802676;  
 RA Sohma M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;  
 RT "Molecular cloning and sequence analysis of a human 372-kDa protein  
 RT localized in the Golgi complex";  
 RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).  
 CC -!- FUNCTION: May participate in forming intercisternal cross-bridges  
 CC of the Golgi complex.  
 CC -!- SUBUNIT: Homodimer; disulfide-linked.  
 CC -!- SUBCELLULAR LOCATION: Golgi; membrane-associated.  
 CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the  
 CC autoimmune disease Sjogren's syndrome.  
 CC -!- SIMILARITY: Belongs to the golgin family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X75304; CAAS3052.1; --  
 CC EMBL; D25542; BAA05025.1; --  
 CC FIR; A56539; A56539.  
 CC PIR; I52300; I52300.  
 CC Genew; HGNC:4429; GOLGB1.  
 CC MIM; 602500; --  
 CC GO; GO:000139; C:Golgi membrane; TAS.  
 CC GO; GO:0005795; C:Golgi stack; TAS.  
 CC GO; GO:0016021; C:Integral to membrane; TAS.  
 CC GO; GO:0007030; P:Golgi organization and biogenesis; TAS.  
 CC Antigen; Coiled coil; Golgi stack; Transmembrane.  
 KW DOMAIN 1 3235 Cytoplasmic (Potential).  
 FT TRANSMEM 3236 3256 Potential.  
 FT DOMAIN 3257 3259 Luminal (Potential).  
 FT DOMAIN 48 593 Coiled coil (Potential).  
 FT DOMAIN 677 1028 Coiled coil (Potential).  
 FT DOMAIN 1062 1245 Coiled coil (Potential).

FT DOMAIN 1301 1779 Coiled coil (Potential).  
 FT DOMAIN 1828 3185 Coiled coil (Potential).  
 FT DOMAIN 2420 2423 Poly-Glu.  
 FT DOMAIN 2993 Poly-Ser.  
 FT CONFLICT 1 39 Missing (in Ref. 3).  
 FT CONFLICT 215 215 A -> AQLSSM (in Ref. 3).  
 FT CONFLICT 1765 1765 D -> G (in Ref. 3).  
 FT CONFLICT 2950 2950 H -> D (in Ref. 3).  
 SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;  
 Query Match 4.2%; Score 214.5; DB 1; Length 3259;  
 Best Local Similarity 19.8%; Pred. No. 1.6;  
 Matches 212; Conservative 168; Mismatches 427; Indels 263; Gaps 45;  
 QY 20 HNOEVFSLVKEPIIKQTQASSISGADYAESGSKKLKI-----NETSGPVDVTVDL 72  
 DB 233 HEDELLQLV-----TQAD-----VETEMQKRLVLRKLEHEESLVGRAQVVDL 277  
 QY 73 FSDKRTTPPEKIDNLAKEPREQELK-----AVTENTESEKQITSGSQEQSKESLSLNK 126  
 DB 278 LQELTAARQNRQILSQQLQQMEAEHNTLRNVTETERESKILLEKMELEVAERKLSPH- 336  
 QY 127 TVPSTSNWEICDFITKNTLVGL-----SKSGVEKLSQTDHLVLPQAADGTOLIQV 178  
 DB 337 NLQEEHMLLEQFEQAGQAQAELESRYSALEQKHAEMEETSHLSLOKTG---QELQS 393  
 QY 179 ASFAFTPPDKTAIAEVTSRAGNGEI-----SOLDVDGKEIINEGEVFNLSLLKVTIPT 233  
 DB 394 ACDALKQNSKLLQDKNEQAVOSQTIQOLEQLOQSKKEI-----SQFLNRILPQQ 445  
 QY 234 GYKHIGQDAFVDNKNIAEVLNLPESLETISDYAPAHALKQIDLPNKLKAIGELAPFDNOI 293  
 DB 446 -HETASQTSFPDYN-----EGTQAVTEENIASLQKRVVEL-----ENEK 484  
 QY 294 TGKLSLPQLMRL-AERAPKSNHIIKTIEFRGNSLVKIGEASQDNDLSQMLPDGLEKTE 352  
 DB 485 GALLLSISIEBELKAENEKLSQITLLEAQNRT---GEA---DREVSEISIVDIANKGS 537  
 QY 353 SEAFPTGNPCDDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLWQESPIDYTKWLEEDF 412  
 DB 538 SSA-----BESGD-----VLENTFSQKHKL-----SVLLEKKAQEERI 573  
 QY 413 TYQKNSTVG-----FSNKGLOKVGKRNKLE-----IPKQNGVT 446  
 DB 574 AFLKLQLOQKRAEADHEVLQKEMKQMEGEGIAPIKMKVFLDTQDQPLMPNESSLIP 633  
 QY 447 ITEIGDNAFRNVDFQNTLRKYDLBE--VKLPSTTRKIGAPAFQSNLNKSFASDDLEI 504  
 DB 634 AVE---KEQASTEHQSRSTSEISLNDAGVELKST-----KQDGKSLSAVDPDQGC 681  
 QY 505 KEGAPMNNRIETLELKDKLVTIGDAAFH-INHIYAILVLPESVQETCR-----SAFRQGA 558  
 DB 682 HODELERLKSQLEL-----ELNFHKAQEIYEKNLDEKAKEISNLNQLIEEPFKKAD 733  
 QY 559 NN-----LIFMGSKVKTIGEMAFISNLEHLDL-----SEQQLTEIPVQAFS 601  
 DB 734 NNSAFTALSEERDQLL---SOVKELSMVTELRAQVKLEMLNLAERQRLDYESTQTAH 790  
 QY 602 DNALKEVL---LPASLKTIREAPKFNHKLKOLEVASALSHIAFNALDNDGDGEQDNK- 656  
 DB 791 DNLLTEQIHSLSIEAKSKDVKIEVL-QNELDDVQLQFSEQSTLIRSL-----QSQLNQKE 844  
 QY 657 -----VVVYKTHNSVALADGEHFIVDPKLSLSTIVDLAKILKLEGLDYSTLR 704  
 DB 845 SEVLGAERVRHISSKVEELSQALESLEITKMDOLL-----LEK-----KRDVETLQ 893  
 QY 705 QTTOFTQFDMTTPAGKALLSKSNLRQGEKQKQFQEAQFPLGRVLDLQKAIKAELVTKKA 764  
 DB 894 QTIEEKDQVVT-----EISFSTWTEKXVQLNEEFSLG-VET-KTLKEQLNLSRAEE 943  
 QY 765 TKNGQLLERSINKAVLAVNNSAIKKA---NVKRLKEKELDLLTGLVEGKGPLAQATWQGV 821  
 DB 944 AKKEQVEEDNEVSSGLKQNYDMSAPGOISKEELQHEFDLLKKENEQRKRKLQAALINRK 1003

```
QY 822 YLLKTPPLPEYIYGLNVYFDKSGKLIYALDMSDT-IGEGQKDAYGNPILNVDEBNEGYH 880
Db 1004 EQLQVRSLBELANLK---DESKK---EIPLSETERGEVEE-----DKENKEYS 1047
QY 881 ALAVATLADYEGDLDKTILNSKLSQTSIR-----QVPTAAVHRAGI 922
Db 1048 EKVTSKQCEIYILKQTISEKEVQLQIRKDLKLAABEQFQALVKQMNTLQDKTNQ 1107
QY 923 FOAIQNAAAEAQLPKPCTHSEKSSSES-ANSKDRGLQSNPKTNQRH 971
Db 1108 IDLLQAEISENQAIQKLTISNTDASDGSVALVKETVVISPPCTGSSEH 1157

RESULT 35
Q97NB5 PRELIMINARY; PRT; 621 AA.
AC Q97NB5;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DE Choline binding protein PCpA.
GN OrderedLocusNames=SP2136;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.B., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AE007502; AAK76194.1; -.
DR PIR; A95250; A95250.
DR HSP; F06653; IH86.
DR TIGR; SP2136; -.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR007093; LRR Tp.
DR Pfam; PF01473; CW_binding_1; 7.
KW Complete proteome.
SQ SEQUENCE 621 AA; 69237 MW; 69808PDP43895EB6 CRC64;

Query Match 4.2%; Score 213.5; DB 2; Length 621;
Best Local Similarity 23.2%; Pred. No. 0.18;
Matches 132; Conservative 75; Mismatches 175; Indels 187; Gaps 31;

QY 6 KTVALTLTIVSVV-----THNQEVF-----SLVKEPILK-----Q 35
Db 3 KTTILSLTAAILAAYVNPBILADTPSEVIKTKVGIISQNNIKYKLVTVEGNIGT 62
QY 36 TQASSISGADY-ARESSGK-----SKLINETSGVDVDTVDLFS-----DK 76
Db 63 VOVGNGVTPVFEAGQDQKPFTPIKTVGOKVFTTVEVAGQFSYYPDETGRIVYPPSS 122
QY 77 RTTPPEKIDNLAQPREQELKAVTENTESEKQITSGSQLEQSK-----120
Db 123 ITIPSSIKKIQKGFHSGKAKTII-----FDKGSQLEKIEDRAPFSELEIELPAS 174
QY 121 -----SLSLANK-TVPESTSNWBIKCDFTKNTLVGLSKSGVEKLSQTDHLVLPSQ 168
Db 175 LEYIGTSAFSFSQKLKLTFFSSSKLEL-----ISHEAFANLSNLEKLTLPKS 222
QY 169 AAD-QTQLIQVA-----SFA-----PTPKKTAIAEYTSRAGENGISQLD 208
```

```
Db 223 VKTLGSLNLFRLTTSKLVNDVEEGNESFASVDGVLFSD-KTQLIYYPSQ--KNDESXTP 279
QY 209 VDGKEIINBEGEVFNYSLLKKVVTPTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAF-A 267
Db 280 KETELASYSFNKSY-LKKLELNEGLEKIGTFAFADAIAKLEESLNSLETIERLAFYG 338
QY 268 HLAUKQIDLDNLKAIQELAP-----FDNQITGKLSLPRQLMRLAERAFKSHIHK 317
Db 339 NLELKEILFDNVKNPGKVMNGLPKLSLTIGNNIN---SLPSFFLSGLVDSLKEIHK 395
QY 318 -----TISFRGNS--LKVTGEASFQ-DNDLSQLMLPDGLEKTESBAFT 357
Db 396 NKSTEPSVKQDTPAIPETVFPYVTSSEHKVKNLSLSTNDI-----IVEKVDNLIKQETDV 451
QY 358 GNPDDHNNRNVVLWTK-----SGKNPSGLATENTYVNPDKSLW---QBSPRIDY 404
Db 452 AKP-KKNSQGVGVGWKDKGLWYLYNESGSMATGWV-----KDKGLWYLYNESGSM-A 502
QY 405 TKWLEED--FTYQKNS---VTGP-SNKG 427
Db 503 TGMVKDKGLWYLYNESGSMATGWYKDKGL 531

RESULT 36
Q71WB3 PRELIMINARY; PRT; 489 AA.
AC Q71WB3;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE Cell wall surface anchor family protein.
GN OrderedLocusNames=LWof2365_2638;
OS Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=265669;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15115901; DOI=10.1093/nar/gkh562;
RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,
RA Kolonay J.F., Rasko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,
RA Peterson J.D., White O., Nelson W.C., Nierman W.C., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,
RA Haft D.H., Selengut J., Van Aken S.E., Khouri H.M., Fedorova N.,
RA Forberger H.A., Tran B., Kathariou S., Wonderling L.D., Uhlrich G.A.,
RA Bayles D.O., Luchansky J.B., Fraser C.M.;
RT "Whole genome comparisons of serotype 4b and 1/2a strains of the food-
RT borne pathogen Listeria monocytogenes reveal new insights into the
RT core genome components of this species."
RL Nucleic Acids Res. 32:2386-2395(2004).
CC -|- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (by similarity).
DR EMBL; AE017331; AAT05403.1; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR InterPro; IPR009459; DUF1085.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR007093; LRR Tp.
DR Pfam; PF06458; DUF1085; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Complete proteome; Peptidoglycan-anchor.
SQ SEQUENCE 489 AA; 53529 MW; 07FA798B2A3B86 CRC64;

Query Match 4.2%; Score 213; DB 2; Length 489;
Best Local Similarity 21.0%; Pred. No. 0.13;
Matches 118; Conservative 87; Mismatches 193; Indels 164; Gaps 21;

QY 472 EVKLPTSTIKIGAPAFQSNLKSFEASDDELEIKEGAPMNNRIETLBKDKLVITGDAAF 531
Db 51 DITITPTL-----GPNNEYTVTA-----IGNGAFKSKRLTNVITPTVITIGDAF 96
```





Db 1632 DQLEAEQQHVABLE 1645

RESULT 38

Q7RGP2 PRELIMINARY; PRT; 1850 AA.

ID Q7RGP2

AC Q7RGP2; (T-EMBLrel. 26, Created)

DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)

DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)

DE Hypothetical protein.

GN Name=PY04304;

OS Plasmodium yoelii yoelii.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

NCBI\_TaxID=73239;

[1]

SEQUENCE FROM N.A.

RC STRAIN=17XNL;

RX PubMed=12368865; DOI=10.1038/nature01099;

RA Carlton J.M., Angluoli S.V., Suh B.B., Kooij T.W., Pertea M., Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L., Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L., Shallom S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V., Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M., Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A., Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B., van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R., Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J., Carucci D.J.;

RA "Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii.";

RL Nature 419:512-519(2002).

CC -I- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBO whole genome shotgun (WGS) entry which is preliminary data.

DR EMBL; AABL01001295; EAA16146.1; --

KW Hypothetical protein.

SQ SEQUENCE 1850 AA; 215313 MW; 00366A8B5CE21629 CRC64;

Query Match 4.2%; Score 213; DB 2; Length 1850;

Best Local Similarity 19.1%; Pred. No. 0.84;

Matches 204; Conservative 175; Mismatches 364; Indels 324; Gaps 51;

QY 6 KTVALLTVSVVTHNQEVSLVKSPILK-----QTQASSISGADYAESSCKSLKI 58

DB 489 RRVSVTSKYNFKITEEENDMLKRDKLSNKSVSRRKTNANGLDILEKVTENK-NI 547

QY 59 NETSGPVDDTV--TDLFSKRTTPKIKDNLAKGPREQELKAVTENTSESEKQITSGSLE 116

DB 548 NNNDVSFTDKIFSCNNFNIMT--TFKDNLFKEPKYSKLDHNEINFEKYK----GDMLY 601

QY 117 QS---KESLSLTKTVSTSNWEICDFITKNTLVGLSKSGVEKLSQTDHLVLPQAA--- 170

DB 602 YSINDKNGINENIGLSKN-----ITFFNT---NEDVDQIDRIDTIDPLNKLRLSMYG 651

QY 171 -----DGTQLIQ-----VASFAFTPDKTAIAEYTSRAGENGEL----- 204

DB 652 KSYDDIINEGNINNNNDKNINYNPTLGLSMKFPINSSVMINQWQFDGDDINNIPINRSK 711

QY 205 -----SOLDV-----DGKEIIN-EGEVFN-----SYLLKKVT 230

DB 712 SYTNIIIGNSNDIYKNTNNINNNKFNMSNDPKSGYQIVTEGENINFRNSSVESKIRN 771

QY 231 IPTGYKHIGODAFVDNKN--IAEVN-LPESLETISDYAFHALKQ-----IDLDPNL 280

DB 772 IE---KYIN-----DNKNNAIRKSNYINKYMTHTDNTFGGDTINKMNTNFHIDNPILL 823

QY 281 KAIGELAFDNIQTGLKSLPQRLMLRAEPKSNHIKTIIEFRGNSLKVIGESAFQDNDLS 340

DB 824 -----INNINRIM-----NSVKGDINKVE---NYEQWTSQEPFKQDIT 862

QY 341 QLMPLDGLKIESEAFTPNGDDHYNNRV--LWTKSGKNPSGLATENT--VYNPKSL--- 395



TIGR; EF1824; .  
DR GO:0009986; C:cell surface; IEA.  
DR GO:0016787; F:hydrolase activity; IEA.  
DR GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO:0000715; P:cell adhesion; IEA.  
DR GO:0000272; P:polysaccharide catabolism; IEA.  
DR InterPro: IPR008965; Cellul bind.  
DR InterPro: IPR002105; Dockerin\_1.  
DR InterPro: IPR002048; EF-hand.  
DR InterPro: IPR000421; FAS8 C.  
DR InterPro: IPR011490; FIVAR.  
DR InterPro: IPR003961; FN III.  
DR InterPro: IPR000322; Glyco hydro 31.  
DR InterPro: IPR001899; Gram pos anchor.  
DR Pfam: PF00754; F5\_P8\_type\_C; 1.  
DR Pfam: PF07554; FIVAR; 6.  
DR Pfam: PF00041; fn3; 1.  
DR Pfam: PF01055; Glyco hydro 31; 1.  
DR SMART; SM00060; FN3; 1.  
DR PROSITE; PS00448; CLOS CELLULOSOME RPT; 1.  
DR PROSITE; PS00018; EF HAND; UNKNOWN\_1.  
DR PROSITE; PS00022; FAS6C\_3; 1.  
DR PROSITE; PS00853; FN3; 1.  
DR PROSITE; PS00847; GRAM POS ANCHORING; 1.  
KW Complete proteome; Hydrolase.  
SQ SEQUENCE 1866 AA; 205565 MW; 4682D5E8B877A96D5 CRC64;  
  
Query Match 4.2%; Score 213; DB 2; Length 1866;  
Best Local Similarity 19.3%; Pred. No. 0.85;  
Matches 239; Conservative 176; Mismatches 423; Indels 398; Gaps 56;  
  
QY 49 ESSGSKLAKINETSQGVDTVTDL-----FSDKRT-----PEKI--- 83  
DB 751 EAQQSATTIN-SQGPKNEKGLTVTIBFTKSYKDFVDERSTTLDLASEASVVTAM 809  
  
QY 84 ----KDLAKPREQELKAVTENTESEKQI-----TSGSOLEQSKESLSINKTVPST 131  
DB 810 VCGTEVTLQOANKKEFLAGTNLYYDFKEFQVNOYLSEASGEKLNQSLSVKLAKQSVTA 869  
  
QY 132 SNEIC--DFITK-----GNTLVGLSKSGVEKLSQTDHLVPSQAA----- 170  
DB 870 KDQVITVKGFINKGTVDGGNTTV-----DDQLTIPANVAINEKTTTPSSLTQ 917  
  
QY 171 -----DGTQLIQVASFAPTPDKKTAIBYTSR---AGENG--EISQ----- 206  
DB 918 WDQVTEATSYEVEDGTQFNQIQTNTATDFGFSFLSEHTFRVRANGKGVSEWSEPIK 977  
  
QY 207 -LDVKGKEIINEGEVFNYS-----LLKKVT---IPTGVKHIGQDAFVNKNIAEVNLP 256  
DB 978 TODDPYKETINQVATSNLPEQGAELKLTDKDLSTGW-----HTNWT 1022  
  
QY 257 SLETTISDYAFHALK-----QID-----LPDNLKAIGEL-----AFFDN 291  
DB 1023 GIAPNSDGNF--LSLKFDLGAEQMDKIEVLPDNGAGNLIQLQVYRTSKDGNWTFESE 1080  
  
QY 292 QITGKLSLPQMLRLAERAFKSNHKTIEFRGN-----SLKVIGEA 332  
DB 1081 PINWKODALTKTITKQDQYRFVEMKVLKSVGNFGSGREMLFYKQPGTEGILHGDITNDG 1140  
  
QY 333 SFQDNDLSQLMPLDGLKTESAFTG--NPGD-----DHNNRVVLWTKSG----- 376  
DB 1141 TIDENDAMSYRNTGLSDSD--FNGYVEKGDUNKNGVIDAVDISVVLQDGGIETIPDV 1199  
  
QY 377 -----KNPSGLATENTYVNPDKSLWQESPE 401  
DB 1200 EETAGLSLAVNNGKDTYLPDGTTLTKGDLKLNALSTKWSFDSKSELVGPAT 1259  
  
QY 402 IDYTKWLEDDFYQKNS-----VTGFSNKGKQKVRKNRLKIEPKOHNGVTITGDNFA 455  
DB 1260 TNNQTQOMENYSKYRKHSDVENLYLVLSNQ-----NKQL-----LNG----- 1297  
  
QY 456 RNVDQFQNTLRKYDLEEVKLPSTIRKIGAP-----AFQSNNLKSFASDDLEEK 505

DB 1298 -SMDLVTFKVKVKTTRVRKATTVTQPLQFDMSQGLLVGQGFQATLSDFSVT----- 1349  
QY 506 EGAFNNRIETLELKK-----LVTIGDAAFHINH-----IYAVLPSPVQEIHSAPR 554  
DB 1350 -----VKPTLVDKELLQALITLNQARVEKYTPTWAFKFPILDEAVLANEQAT 1401  
QY 555 QN-----GANNLIFMGSKVKTLGCEAFNLSRLHDLSEKQOLTEIPV--QAFSDNALK-- 606  
DB 1402 QTDVSAAAEENLEKAASOLEKMPDVA--NKAD-LEKATQEGGLAKPSCQGFTEETKKVL 1457  
QY 607 EVLLPASLKTITREAFKKNHLKQLEVASALSASHIAFNALDDNDGDEQFQDNKVVVKTTHNSY 666  
DB 1458 EESLAAAKVFAQEKVTOEIDQ-----ATKTLREAI 1489  
QY 667 ALADGEHFIVDPDKLSSTIVDLEKILKLEGLDYSITLQTTQTFQFRDMTTAGALLSKSN 726  
DB 1490 AOLKEQPVAVDKETLKEQIAQ--ARGKPEEGYQFT---KETEKQLQEAIQAAEAIVAKET 1545  
QY 727 LRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKA-----TKNGQLLERS---INKA 778  
DB 1546 ATKEEVSEALNALETAMAKQLEKVPVNVKQDQLEVVVRAQQVTPSEGHQFTASSLOELQA 1605  
QY 779 VLAYNNSAIKK--ANVXLEKEKELDLTLGLVEGKGLAQATMVQGYLLKTPPLPEY-- 834  
DB 1606 LLAAGNT-LKNPAANQKIMDEAVELTSAIDG---LQEBVLVTDKKALEAMIAKAKIKP 1661  
QY 835 -IGLVNVPDKSLIYALDMSDTIGEGQKDAYGNPILNVDENEGHEHALAVATLADYEG 893  
DB 1662 SAGKEFTSESKARLTEAIDQAEGL-LADKNARQEQI-DIAEKN-----VKTALD--SL 1710  
QY 894 DIKTILNSKLSLTSIROVPTAAVHRAG--IQAIONAAE-----AEQLLPKPGTHSEKS 947  
DB 1711 E-EQVLQDQTKLKELOKAEETLKPKAGKQFTKASQEALEAIKQAKALVEDPNATQEA 1769  
QY 948 SSSSEANSK-----DRGLQSNPKNTRGRHSAI----- 974  
DB 1770 DKCLSILQAEAMAEPEISSNNGNHSITVSGTSGKGTATGTTTITTSCT 1829  
QY 975 LPRTSGKSGFVVGILGY---TSVALLSLITAIKXKK 1007  
DB 1830 LPKANEVVSPISWSISGFLLIIVSIGLGLFFKNKKEQ 1865  
  
RESULT 40  
Q9AOK5 PRELIMINARY; PRT; 2045 AA.  
AC Q9AOK5;  
DT 01-JUN-2001 (TREMELrel. 17, Created)  
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Putative extracellular matrix binding protein.  
GN Name=epf; OrderedlocusNames=SPY0737;  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;  
RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;  
RA Perretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
RA Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,  
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;  
RT "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).  
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
an amide bond (By similarity).  
EMBL; AE006525; AAK33683.1; -.  
DR GO:0009986; C:cell surface; IEA.  
DR GO:0005618; C:cell wall; IEA.  
DR InterPro; IPR011439; DUF1542.





Db 1823 -----EGLLKLEDTLRN-----IKLTGTGAPFASL-----SELEKDLIS 1858  
QY 472 EVKL-----PSTIRKIGAFAPQNNILKSPASDLBEI-----KEGAFPMNRIETLELKDKL 523  
Db 1859 ENRLAAVDEASVRGIG-----KLREFYINDN-----EIGIDGLSGLDILNRVQTVSLSNQ- 1908  
QY 524 VTIGDAAAPHINHIYAIVLPESVOETGRSAPFRONGANNILFMGSKVKTIGEMAFSLN-RL 582  
Db 1909 -----SIVPDTL-----LANKSMVFYVQGTTPASLPAOFFRSTIAL 1948  
QY 583 HLDLSEKQLTEIPVOAF-----LPA-----SLKTIREAFKKNHLK 607  
Db 1949 KLSLSDNQLVKLPDQFQYVPLBELVLLNNSLSVLPFGVDPGLIRELDISENPLKS 2008  
QY 608 V-----LPA-----SLKTIREAFKKNHLK 628  
Db 2009 LPGLGFAQTFLLESRLMADANTLTPAGIFDKLYVLAKLDLANNQLRTLREGVF--NRLY 2066  
QY 629 QLEVAS-----ALSHAFNALDNDGDGEQDNKVVVKTTHNSVALADGEHFIVDPDKL 681  
Db 2067 SLETLSLENNQLEALQALFKSLEKLN-----IVLSHKNLAADPQLFAGLP--- 2114  
QY 682 SSTIVDLEKILKLEGLDYSYLRQTQTFQFRDMTTAGKAL-----LSKSNLRQGEKQ 733  
Db 2115 -----VTAIDLSYNRLTTFDE--SAMAFAGRLMLSMESNKILTSKIHSTLEK 2160  
QY 734 KFLQBAQFPLGRVLDLKAIAKAE-----KALVTKKATKNG--QL-LE 772  
Db 2161 LPVDDNKLSSIEVKLEDSYSHITSLSVNNTFTALDAFYQFQMLTELDASYNAFTELNLA 2220  
QY 773 RSINKAVLAY-----NNSAIKKANVRLEK-----ELDLTLGLVEGKGLAQTWVGVL 823  
Db 2221 KLFNEIVPLYTLNVNSFVERVQFGKIEHTPTLVHLDISNNMLTALDVRAFAVFRQ----- 2276  
QY 824 LKTLPLPEYTYGLNVPYDK--SGKLIIVALDMSDTIG--EG---QKDAYGNPILNDE--- 874  
Db 2277 -----LQNFVPGGN--REDTFVEVDLLRAFDDLEITIGLEGTVMKCDFLKSLSKASMEADV 2329  
QY 875 -----DNEGYH--ALAVATLADYEGDLIKTLNLSKLSQLT-----SIRQVPTAAYHRAGIF 923  
Db 2330 TPAYKEKETLHESGLRFATPSTASLSI--TFPSDYVSHVTPENSLSAIPQTVPAKSGV 2388  
QY 924 QAIQNAABAEQLLPKPGTHSEKSSSSSANGKRGSLGNPKTNRGRHSAILPRTGSKGS 983  
Db 2389 EFTNASGCDIYELLPKTRFH-----AADLKLYLDNNR-----LRKT----- 2425  
QY 984 FVYG--ILGYTSVALLSL 999  
Db 2426 --YGSNFFLGARSLELLSL 2441

## RESULT 43

Q98QZ9  
ID Q98QZ9 PRELIMINARY; PRT; 3216 AA.  
AC Q98QZ9  
DT 01-OCT-2001 (TremBLrel. 18, Created)  
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)  
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)  
DE Hypothetical protein MYPU\_2110.  
GN HyperdocusNames-MYPU\_2110;  
OS Mycoplasma pulmonis.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxId=2107;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=UAB CTIP.  
RA MEDLINE=21267165; PubMed=11353084; DOI=10.1093/nar/29.10.2145;  
RX Chandaud I., Heilig R., Ferris S., Barbe V., Samson D., Gallison F.,  
RA Mosser I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,  
RA Blanchard A.;  
RT "The complete genome sequence of the murine respiratory pathogen  
Mycoplasma pulmonis."

RL Nucleic Acids Res. 29:2145-2153 (2001).  
DR EMBL; AL445563; CAC13384.1; --.  
DR PIR; C90538; C90538.  
DR MyPUList; MYPU\_2110; --.  
DR InterPro; IPR007326; Lipoprotein 17.  
DR Pfam; PF04200; Lipoprotein 17; 5.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 3216 AA; 359419 MW; A62ED957568DB90C CRC64;  
Query Match 4.2%; Score 212.5; DB 2; Length 3216;  
Best Local Similarity 19.9%; Pred. No. 1.9;  
Matches 218; Conservative 148; Mismatches 380; Indels 347; Gaps 55;  
QY 7 TVALTLTT--VSVVTHNQEVPSLVK-----EPILKQTAQASSISGADYAESGGSKLKLIN 59  
Db 2274 TLTITISSDQTSRTTYQETIEGFVSDNRAENIYAVKGYLSEIQASGRAPSGTGTVI- 2332  
QY 60 ETSQFVDD-----TVTDLPSDKRTTPE--KIKONLAKGPREQELKAVTENTSEKQITS 111  
Db 2333 ----PILDPSVIQNNSTDEFVNKILTNHASLGININKNPQ-----SNNSNLEQSLIS 2381  
QY 112 GSQI---EQSKESISLNTKTPSTSN-----WEI--CDFITK--GNTLVGLSKSGVEK 156  
Db 2382 AEVLTTNEYPAESVKLIQVKSKKGDQDHYSRVYVYSGFIQRTGADILNIRRAIBAS 2441  
QY 157 LSQTD---HLV-----LPSQAADGTQLIQVASFAFTPK-KTAAEYTSRAGE 200  
Db 2442 LRENDKAQHIKVKKESREKITKTISTTTVGKNFTTSDFEINNELIKTKVTNVQINDY 2501  
QY 201 NGEI--SOLDVDGKEIIEGEVFNYSLLKQVITPTGYKHIGQDAFVDN--KNIAEAVNLE 256  
Db 2502 NVELSLVDJISRENIISGNARVIRISK-----DNSEKSEFIVNI--S 2544  
QY 257 SLETISDYAFALAKQIDLPD-----NLKAIAGELAPF---DNQITGLSLPRQL 303  
Db 2545 GFETLSEAFDKYIKSSTRFIPSVKEANKOLWDTKATSENFSTPSQNKINDKPDFL--- 2600  
QY 304 MRLAER-----AFKSNHIIKTIETPRGNSLKVIGEASFQDNDLSQMLPDGLEKIES 353  
Db 2601 -AVKERFVFSVDTSKVFKMELSKTIKTINFGISSTFQSAQYEDV-----DGFKK--- 2650  
QY 354 EAFTGNPDGDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKMLEEDFT 413  
Db 2651 ELFDTN---DIFR--VOYEKQK-----TINKYVKE--KNSHNIDFTK----- 2688  
QY 414 YQKNSVTGFSNGLQKVRKNKLEIPKQHNQVTTITEIGDNAPRVDFQNTLRKYDLEEV 473  
Db 2689 -----IKKLEI-----GPNFSEIPKDYFKDAS-----SLTEL 2715  
QY 474 KLPSTIRKIGAFQSNILKSFASDDLEIEKEGAFPMNRIETLELKDKLVTIGDAAPHI 533  
Db 2716 KINSQVSKIKESAFESAKLTSLFNSLVEIGPNAFKNSVLTSL----- 2759  
QY 534 NHIYALVPESVOETGRSAPFRONGANNILF--MGSKVKTIGEMAFSLNRLHLDLSEKQ 591  
Db 2760 -----SGLEQTKLSALKEN-----VPEKANDSIKTT--IWNKKELAKEPI 2800  
QY 592 LTEIPVQAFSDNALKEVLLPASLKTIRBEAFKQKHLKQLEVASALSHIAFNALDNDGDGE 651  
Db 2801 LT-----STNK-----IPSED--KKAKKPSQFEQ-----SNFSPAATKQEDSDI 2838  
QY 652 QPD-----NKVVVKTTHNSVALADGEHFIVDPDKLSSTIVDL-----EK 690  
Db 2839 HFEITGFSQDDVFGFTIETTYKIKTKTDESFPVFDEN-----NPSK--KVLDTFTNLEEK 2892  
QY 691 ILKLEIGLDY-----STLRQTTQTPFRDMTTAGK--ALLSKNLRQGEKQFLQEAQF 741  
Db 2893 VTKIKDASSYFDTSSKIETIEKTKDTSNNTSTENQNTKRLVQWKDQVNGKSGSVFLPQSSV 2952  
QY 742 ---FLGRVDL--DKAIAKAEKALVTIKATPKNQQLLE--RSINKAVLAYNN----- 784  
Db 2953 NAPPVETTTQIGDKTTSLVDLDFNNPQSQIRNNLSFENLKPSSAVLWNSDASLKNQSKIT 3012



RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
 RA Richardson P., Rubin E., Rice H.;  
 RT "Complete genome sequence of *Bacillus thuringiensis* 97-27.";  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
 an amide bond (By similarity).  
 DR EMBL; AE017355; AAT61022.1; -;  
 DR GO; GO:0009986; C:cell surface; IEA.  
 DR GO; GO:0005727; C:extrachromosomal circular DNA; IEA.  
 DR InterPro; IPR001434; DUF11.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR Pfam; PF01345; DUF11; 4.  
 DR TIGRFAMs; TIGR01451; B\_ant\_repeat; 19.  
 DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.  
 DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
 KW Cell wall; Complete proteome; Hypothetical protein;  
 KW Peptidoglycan-anchor.  
 SQ SEQUENCE 3471 AA; 386275 MW; C27F7C668311ABA7 CRC64;

Query Match 4.2%; Score 212; DB 2; Length 3471;  
 Best Local Similarity 19.8%; Pred. NO. 2.2;  
 Matches 215; Conservative 166; Mismatches 362; Indels 342; Gaps 57;

QY 44 GADYAESGSKLKNINETSIPVDDTV-----TDLFSDKRTTPEKIKONLAKGP 91  
 DB 1225 GLEYVENSIAKAE---GEAPSPVELTVAGVVKAKYIDITDKERSIVFKVKVEVEVG- 1280

QY 92 RQELKAVTENT-----ESEKQITSGQLEOSKES-LSLNKTV--PSTSNWEICDF-ITK 142  
 DB 1281 KEIVNKAIVDDTKHQPIPERIT-----PQKQGIINAAKIVDNPSPKLGEEVEYRISF 1335

QY 143 QNTLVGLSKSGVEKLSQTDHL-----VLPQADGTQLIQV-----ASFAPPD 186  
 DB 1336 KNT---VENGKLEKVKIEDTTPNGLEYVKGSEKAGDKPAPVKLVHVGKGVIAEYENITD 1392

QY 187 KKAIAEYTSRAGENETISQLDVGKEIINEGEVENSVLAKKVTIPTGYKH----- 237  
 DB 1393 TKERSIVTVKVEAEI-----GKEIVNQAIVDVTKDSKPEAKITPLHKOGKIKAKK 1446

QY 238 -----IGQDA-----FVDNKNIAEV-----NLPSESLETISDYAFAPHLALKQIDLP 277  
 DB 1447 SVNNETPKLGEVEYRISFKNTVENGKLAEVKIEDTLPGLGYVE----- 1491

QY 278 DNLKAIG-----ELAFDQNIQTKGLSLPQMLRAEA--PKSNHIKTIIFRGNLSKVI 329  
 DB 1492 NSLKAEGAGNPVELKMGKVLAKY---ETTDIERSITPK---VKVSEVKVGEKIV 1545

QY 330 GEASQD-----NDLSQL--MLPGLKIESEAFTHNGP---DDHYNNRVVLTKSGKN 378  
 DB 1546 NKAIVDDTKHDPIPKASITPQYKDG--KIEAEKVNNPSPKLGEEVEYRI-----SPKN 1598

QY 379 -----PSG-----LATENTVYVNPDKSLMOES-----PEIDYTKWLE 409  
 DB 1599 TVEHGKLAEVIEDLPNGLEYVKSQLEAGSKPNPVELKPKENGKVLAKYFEITDK--E 1656

QY 410 EDFTYQ---KNSVTG--FSNKGIL-----QKVRNKN--LEIPKQHNGVTITEI 450  
 DB 1657 RSITPKVKVKNVSDTIINQAIVSDTKHPPEPKAEIIPQHKDGKLEAKVNNL-LPKL 1715

QY 451 GDNAPRVDPQNTLRKYDLBEVKLPSTIRKIGAFAPQSNLAKSFEASDDLEEIKGAFM 510  
 DB 1716 GEEVEYRISFKN--TLENGKLAEVKIEDTLP--GLEYVENSILKAEAGTDSVELK---FE 1769

QY 511 NNRI-----ETLELKDVLVTIGDAAFHINHIYAIVLPSVSQVIGRSAPRQNGANNLI 565  
 DB 1770 NGKILAKYPEITDKERSIT-----PKVKVDEVIKGGI-----VNKAIDDD 1812

QY 566 SKVK-----TLGEMAFNLRLBHLDLSEKQQLTETIPV 597  
 DB 1813 TKKEPETTAETIPQHKDGKVEAEKTVNNPSPKLGEE--IEYRISFKNTVENGKLAEVKI 1870

QY 598 QAFSDNALKEVLLPASLTIREEAPFKNHLKQLEVASLSHLSHAFNALDNDGDGEQFDNKV 657

Db 1871 EDTLPNGLEVY-----KDSIQAGSKPNPV-ELKVGKQVIKAKYPAITDIE-----ERSI 1919

QY 658 VVKTHNSYALADGEHFIVDPDKLSSITVDLEKILKLEIGLDYSTLRQTQTOTQFRDMTTA 717

Db 1920 VFKAQV-----EDFKVGSIVKVVVDTK-----DPKTREVTVTPEYKD--- 1960

QY 718 GKALLSKSNLRQGEKQKFLQEAQFFLQVRVLDLDKAIKAEKALVTKKATKQGLLERSINK 777

Db 1961 GK--LKAEEKFVNKKPKLGEVEY---RINF-----KNTVENGLKLEVEVKVED 2002

QY 778 AV---LAYNNSAIKANVKRLEKLELDLTGLVEGKGLAQATMVQGVVL---LKTPLPLP 831

Db 2003 EIPAGLSYVNSLQAGSKPSPVELKPFENGKVMKAPVITDKERSIVFKKAKVEEABIG 2062

QY 832 EYITGLNVYFD-----KSGKLIYALDMSDTTIGEGQKDAYGNPILNVDED 875

Db 2063 KEIVNKAIVDDTTHEPEKPYVEITPQYKDGKIV-----AEKVANNHKPKLGEV--- 2111

QY 876 NEGHYALAVATLADYEGLDIKTILNSKLSQTSITRQVPTAAHYHRAGIFQAIQNAAEABQ 935

Db 2112 -----BYRIRFKNTVESGKLAEVNIKDTLPKGLGYVEG-----SITAEG 2150

QY 936 LLPKP 940

Db 2151 SKPKP 2155

RESULT 46

Q6BY65 PRELIMINARY; PRT; 1133 AA.

AC Q6BY65;

DT 25-OCT-2004 (TrEMBLrel. 28, Created)

DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)

DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)

DE Debaryomyces hansenii chromosome A of strain CBS767 of Debaryomyces hansenii.

GN ORFNames=DEHA0A12507g;

OS Debaryomyces hansenii CBS767.

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.

OX NCBI\_TaxID=284592;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CBS767;

RG Genolevures;

RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Tallia E., Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykaeten C., Boisarame A., Boyer J., Catolico L., Confaniolero F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Koszul R., Lemaire M., Lesur J., Ma L., Muller H., Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O., Pellenn S., Pottier S., Richard G.F., Straub M.L., Suleau A., Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierly A., Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J., Winkler P., Souciet J.L.;

RT "Genome evolution in yeasts.";

RL Nature 430:33-44 (2004).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CBS767;

RA Genoscope;

RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; CR382133; CAG84829.1; -;

DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.

DR GO; GO:0006259; P:DNA metabolism; IEA.

DR GO; GO:0000160; P:two-component signal transduction system (p. .); IEA.

DR InterPro; IPR000237; GRIP.

DR InterPro; IPR008207; Hpt.



DR	InterPro: IPR007696; Muts_III.
DR	Pfam: PF01485; GRIP; 1.
DR	PROSITE: PS0913; GRIP; 1.
SQ	SEQUENCE 1133 AA; 130147 MW; 799476682BEE9EC2 CRC64;
	Query Match 4.2%; Score 211.5; DB 2; Length 1133;
	Best Local Similarity 21.0%; Pred. No. 0.5;
	Matches 233; Conservative 156; Mismatches 383; Indels 337; Gaps 54;
QY	30 EPILKQTQASSISGA----DYAESGSKLKLINETSQGVDDVTVDLFSDKRTTPEKID 85
DB	131 EKVLENTPISSISDAGLVLYNG-----VNKTAMLAEMRKYAKENSLSNEKUKD 183
QY	86 NLAKGPREFELKAVTENTSEKOITSGSQLESLSLNKTVTPSTSNWEICDFITKGT 145
DB	184 -----VEARLK-----EALAKDESEGEVTKDSISSQESNDDELKLRN 222
QY	146 LVGLSKSGVEKLSQTDHLVLPQAADCTQLIQVASFAFPDKTAIABYTSRAGENGEIS 205
DB	223 TISDLKGLEK-----KQAEKDT-----VEEGKGENKKEE 252
QY	206 QLD-----VDG--KEIINEGEVFNYSLLKKVTTIPTGVKHIGD--AFVDKNIAEV 252
DB	253 KTDROTKOSEVVDGKYKDIEYKKEIESVKDKLVKSSESRNKHDELSAIERSN----- 308
QY	253 NLPSETISDYAFAHALKQIDLDPNLKAIGELAPFQNQITG-KLSLPROLMRLABRAF 311
DB	309 -----SLEKDKLKDNMSKNDNETIGLTKLEYEKYKQLAEIVD 346
QY	312 KSNHIKT-TEPRNSIKVIGEASPDQNDLSQMLPD-----GLEKISEAFTGNP-GDD 363
DB	347 VNSALETEIENKKEIK-----NFNDISGTQMNDLGANKSTIENLSEAQELNERASD 399
QY	364 HYN-----NRVLWTKSGKNPSGLATENTVNPDKSLMQESPEDITYK----- 406
DB	400 LNLQDERKIKIKELOGLNQSGQPIEA-T-PD-----TPSSNNKKNKKKKGGKKCN 451
QY	407 -MUEEDFTYQKNSVTGFSGNKGLQVRKNKNLEPIQHNGV--TITEIGNDAFRNVDPQNK 463
DB	452 STTDNQIIIVKSEGEQLSAK--PEDIQNKYDELKKENERIOCLKEIDSSKQEISDELKR 509
QY	464 TLAKYP-LBE-----VKLPSTIR-KIGAFQSNLKSF-----EASDLEETK-- 505
DB	510 TKSEYDKLVEEHDDLTHNNHKVSCLKIEDVDHRLDLLTGTDGLVQARDEITKELKAK 569
QY	506 -EGAFMN--RIETLELKDKLVTIGDAAAPHNIHYAIVLPESVQETIGRSAPFGANN 560
DB	570 SKSVLNNDCEUKLETENMQAFPKELESKEIDIE-----KSKQELLKS-IDBNKSSN 620
QY	561 LIFMGSKVKTL---GEMAFSLNRLEHL--DLSEOKQLT-----EIPVOAFSDNALKE 607
DB	621 ----SSNEKLLSTVRGEQKLENFRELTKOLEASEKKNLSLEKEIEKHLXSIEDNLSQ 676
QY	608 VLLPA-----SLKTIREEA-----FKGNHLK-QLEVA 633
DB	677 SLMKADHDALNBKHSALKSKFBETNKQFEIAKQEKXDDLKNIKRIKELSEFKNNDTSLKLEIA 736
QY	634 S-----ALSHT-----AFNALDDNDGDQFONKVVKVTHNSYA 667
DB	737 SLKTSLSHQDLQDIENPKKIIDELKNLBELSGTISKANSELQNSKOLVSEKN---T 793
QY	668 LDAGEHFIVDPD-----KLSSTIVOLEKILLEGIDYSTLR-----QTTOTQ 710
DB	794 LVTKQELSFEKNNSLNNELLKQSEKQKLSYTELTMTKKDHDSLVDKDTSSSNSLOTFRQQ 853
QY	711 FRDMWTAGK-ALLSKNSLRQ--GEKQKFLQEAQFFLGRVDLDJAKAERKALVTWKATKN 767
DB	854 YEELSMKSESLLRDNLDELNELSKTMLQERT--RESSTIRRMVDAEQKLKHNDPFK 911
QY	768 GOLLERSINKAVLAYNNSAIKANVKRLEKDLDTGLVEGKGPLAQATWVGVIYLKTP 827
DB	912 CELTKISEKSEVENNCQAL-----IKKKQREIDEPKSTIDN-----YLLKIQ 954

QY	828 -----LPLPEYYIGLANVYFDKSGKLIYALVMDSDTIGSEGQKDAYGNPILNVDENEGHA 881
DB	955 DLENKCNELKQOYEEL-----QSEKLNKNVTISD-----KDE-----NPEETQQTLET 997
QY	882 LAVATILA-----DYEGLD-----IKTILNSKLSQTSIRQVPPTAAVH--PAGIFQAION 928
DB	998 LRSLQASGSKVKDFENLNLKRLNEESNLKPERLSKNYKILTQOYKQIKSSNDSPIES 1057
QY	929 AAAAEAQLLPKPGT---HSEKSSSSSESAN 954
DB	1058 HAQSSESLTKSPSTSRHSVDDSSSESEN 1086
	RESULT 47
O07910	PRELIMINARY; PRT; 1247 AA.
ID	O07910
AC	O07910; (TREMBLrel. 04, Created)
DT	01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT	05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE	Caga.
DN	Name=caga;
OS	Helicobacter pylori (Campylobacter pylori).
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC	Helicobacteriaceae; Helicobacter.
OX	NCBI_TaxID=210;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 43526, and ATCC43526;
RA	Maeda S., Kanai F., Ogura K., Yoshida H., Ikenoue T., Kawabe T.,
RA	Shiratori Y., Omata M.;
RL	Submitted (APR-1997) to the EMBL/GenBank/DBSJ databases.
DR	'EMBL; AF001357; AAB58747.1; --
DR	EMBL; AB003397; BAA19853.1; --
DR	HSSP; Q9AGB1; IFIM.
DR	GO; GO:0019534; F:toxin transporter activity; IEA.
DR	InterPro; IPR005169; Caga.
DR	InterPro; IPR004355; IVSec_caga.
DR	Pfam; PF03507; Caga; 3.
DR	PRINTS; PR01553; TYPE4SSCAGA.
SQ	SEQUENCE 1247 AA; 138819 MW; 172C76AE9F9BD29E CRC64;
	Query Match 4.2%; Score 211.5; DB 2; Length 1247;
	Best Local Similarity 20.7%; Pred. No. 0.57;
	Matches 226; Conservative 146; Mismatches 389; Indels 329; Gaps 54;
QY	27 LVKEPILKQTOASSISGADYAESGSKLKLINETSQGVDDVTVDLFSDKRTTPEKIKDN 86
DB	1 MTNETIAQQPOTEAFAFPQPFIINNQLVAFK-----VDNAVASYPDQKPVIDK--- 49
QY	87 LAKGPREFELKAVTENTSEKOITSGSQLESLSLNKTV--PSTSNEICDPITKGN 144
DB	50 -----NDRDNQAFDG--ISOURBEYS-NKAIRNPTKKKNQYFDFINKSN 91
QY	145 TL-----VGLSKSGVEKLSQTDHLVLPQAADGTQLIQVASFAF-----TP 185
DB	92 DLINKNDLIDIGSSIKSFQKFQTRQYRIFTSWSHQNDPSKINTSRNFMENIIQPPTP 151
QY	186 DKTAIABYTSRAGEN--GEI--SOLDVDGHKEIINEGEVFNYSLLKKVKTPTGYKHIGOD 241
DB	152 DDKER-ABEFVKSQSFAGIIGNQIRTDQKFM---GVDFELKERQEAENGKEPTGGD 206
QY	242 -----AFVDKNIAEVNLPESLETISDYAFAHALKQIDLDPNLKAIGELAPFNDNOITG 295
DB	207 WLDIPLSFVFNKEQSS-----DVKEAINQESPVPHVQ-----PD-----IATTTHIHQ 249
QY	296 KLSLPROLMRLABRA-----FKSNHIKTITEPRG-----NSLKVIGEASFQDNDSQ 341
DB	250 LPPESEDL--LDERGNFSKFTLGDMEMLDVEGVADIIDPNYKFNQLLI-----HNALSS 301
QY	342 LML--PDGLEKIE-SEAFGTNGP-----DDHNNRVRLWTKSGKNPGLAT 384
DB	302 VLMSGSHNGIEPEKYSLLPAGNGGFCAGKDWATVGYKNQOCDNWATLINVMKNGSGSLVI 361



```
QY 395 ---ENTYVNPDKSLWQESPEIDYTKWLEDFDYQKNSVTGNSKGLQKVRNK--NLBIP 439
Db 362 AGGEGGINP : : : : : SFYLYKEDQLTGSQRALSQBEIRNKIDFMEFL 403
QY 440 KQHNGV--TITEIGNAFRN--VDFQ-----NKTLRKYDLEEVKLPSTIRKIG- 483
Db 404 AQNNAKLNLSEKEKEKQNEIEDFQDSKAYLDALGNDRIAFVSKDQPKPALITEGK 463
QY 484 ---AFAFOSNNLKSFEASDDLEIEKEGAFMNNRIETLELKOKLVITGDAAFHNNIYAI 540
Db 464 GDLSTYLVKMGKKQTEALD-----REKNVTLOGLNKH----- 495
QY 541 LPESVOETGRSFRONGANNLIFMGSVKVTIGEMAFLSNRLHLDLSQKQLTETPVQAP 600
Db 496 --DSVMFVNSYFNKYNASK-----SPDKGVG-----VTNGVSHLDAGFSK-----VAVF 538
QY 601 SDNALKEVLLPASLKTIREEAFFKQKHLKQLEVASALSHIAPNALDDNDGDEQDNKVIV- 659
Db 539 N-----LP-DLNLAIITSFVRNLENKLVTEGLSLOEANKLIK-----DFLSSNKLUG 586
QY 660 KTHNSYALADGEHFIVDPDKLSSTIVDLEKILKLEGLDYSTLRQTQTQFRDMTTAGK 719
Db 587 KALNFKAVADAKN--TGDYDEVKKAQKDLKSLRKEHLE-----K 626
QY 720 ALLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEK-----ALVTKKATKNGQLLERSI 775
Db 627 EVEKLETKSGNKNKHEAKAQ-----ANGQDKIFALINKEANRDARAAYSQ 674
QY 776 N-KAVLAYNNSAIKAN--VRLKELDLTLGLVEGKGLAQATWVGVLKLTPLPLPE 832
Db 675 NUKGIRKRELSDKLEKINTDLKFSKFD-----E 703
QY 833 YYIGLVNVPDKSGKLIYALDMSDTIGEGQKDAYGNP--ILNVDEDEGYH----- 880
Db 704 FKNGKNDKFSKAEETLKALGKS-----VKOLGINPEWISKVENLNAALNRPNGKNKDF 757
QY 881 ALAVATLADYEGLDITKILNSKLS--QLTSIRQVPTAAVHRAGIFAQIAQNAABEQLLPK 939
Db 758 SKVTOAKSLENSVNDVIINQKITDKVDNLQNAVSMKATGG--FGRVEQALADL----- 810
QY 940 POTHSEKSSSSSSANSKORGLQSNPKTNRGRHSAILP--RTGSKGSFV--YGLGYTSVAL 996
Db 811 -----KNFSKEQLAQQ-----TOKNESFNVGKKSSEIYQSVKNGVGTLVNGLSGIEATL 861
QY 997 LSLTAIKKK 1006
Db 862 AKNSDIKKE 871

RESULT 40
Q9AHK6
ID Q9AHK6
AC Q9AHK6
DT 01-JUN-2001 (TreeBLrel. 17, Created)
DT 01-JUN-2001 (TreeBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TreeBLrel. 26, Last annotation update)
DE LMP1.
GN Name=lmp1;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxId=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB19;
RA Dunn J.J.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF305609; AAK18801.1; -.
DR HSRP; F50542; LFCH.
DR InterPro; IPR002713; FF.
DR InterPro; IPR008940; Prenyl trans.
DR InterPro; IPR006597; Sel like.
DR InterPro; IPR001440; TPR.
```

DR InterPro; IPR008941; TPR-like.

DR Pfam; PF01846; FF; 7.

DR Pfam; PF00515; TPR\_1; 4.

DR SMART; SM00671; SEL1; 4.

DR SMART; SM00028; TPR; 9.

DR PROSITE; PS50005; TPR; 6.

DR PROSITE; PS50293; TPR\_REGION; 1.

KW Repeat; TPR repeat.

SEQ SEQUENCE 1173 AA; 134287 MW; F0713ADF2E2311DC CRC64;

#### Query Match

Best Local Similarity 4.1%; Score 210; DB 2; Length 1173;

Matches 224; Conservative 177; Mismatches 393; Indels 404; Gaps 57;

QY 8 VALFTTVSVVTHNOEVFSLVKPEILKQTOASSISGADYAEBSGSKLKINETSGPV-- 65

Db 75 IAMLNTLAEIANSPPES-----KOLQDSANQILDKIKGQDNTKTVNENFDIAFN 127

QY 66 ---DDTVTDLPD-----KRTTPKIKONLAKGPREQ----- 95

Db 128 NRYIKDSTITENYSRNDVDGIEDIEDISEFKKSKIPKIPN--TNPKEBQIIQSNNPK 185

QY 96 -----LKAVENTESEKQITSGSOLEQSKESLSLN----- 125

Db 186 LSVNDQKNLNLKLNKLSGKNSNSENLANSQKIENDKQNTNLSKEKNSENILKTPDNS 245

QY 126 -----KTVPSTSN-----WEICDFITKGNL 146

Db 246 KYSNNNTTSLKIPSNQSELSPPSTQITIGIYRPYSYLIKELYELDDINTGRVT 305

QY 147 VG-----LSKGV-EKLSQTDHLVLPQAADGTLI-----QVASFATPKDKT 189

Db 306 LGRKRLKELIKGLSNKPKVKNELIENSKNEASNLTLTIKKDIEPNLINPKDPYKK- 364

QY 190 AIAEYTSRAGENGISOLDVDGKEIINEGEVFN--YLLKKVTIPTGYKHIGQDAFVD-- 245

Db 365 -----EIFQLDKGDKKQHPGDLKSKVHSIKPIDLENTKSR--QQAIKOLNE 409

QY 246 -----NKNIAEVNLPESLETISDYAFAHALKQIDLPNLK-----AIGELAPF- 289

Db 410 FLKNPNDAQASKTLAQANKIQHLEDLKS RVH---SIKPIDL-ENTKSRQQAIKDLNEFL 465

QY 290 -----DNQITCKLSLPQLMPLAERAPKSNHIKTIEP-----RGNLSKVIGEASPDND- 338

Db 466 KNPNDAAQASKTLAQANKIQHLEDLKS KVSYSIRPIDLENTKSRQQAIKOLNE--FLKNNP 523

QY 339 -----LSQLMLPDGLEKIESEAFNCPGDDHNNRVLVLTWTKSGK-----N 378

Db 524 NDAQASKTLAQANKIQHLEDLKS KVSYSIKPID--LEN-----TKSRQQAIKDLNEFLKN 576

QY 379 PSGLATENTYVNPDK-----SLWQESP-EIDYTKWLEEDFT-----YQKNSVTGFS 423

Db 577 PNDAAQASKTLAQANKIQHLEDLKS KVSYSIRPIDLENTKSRQQAIKDLNEFLKNPNDAQA 636

QY 424 NKGLQKVKENKNALEIPKQHNQVTITECDNAFRNVDFQNK--TLRKVDLEEVK-LPSTIR 480

Db 637 SKTLAQANKIQHLE-----DLSKVSYSIRPIDLENTKSRQQAIAK 675

QY 481 KIGAFAPQSNNLKSFEASDDLEIEKEGAFMNNRIETLE-LKDKLVITGDAAFHNNIYAI 539

Db 676 DLNEFL--KNPNDAQASKTLAQ-----NKIQHLEDLKS-----VHSIRPI 716

QY 540 VL-----PESVQIGRSFRONGANNLIFMGSVKVTIGEMAFLSNRLHLDLSQKQLT 593

Db 717 DLENTKSRQQAIKDL--NEFLKNPNPND-----AQASKTLAQ-----ANKIQHLEDLKS KVSYS 766

QY 594 EIPQAFSDNALKEVLLPASLKTIREEAFFKQKHLKQLEVASALSHIAPNALDDNDGD-- 650

Db 767 IRPIDLENTKSRQQ-----AIKOLNE--FLKNPNDAQASKTLAQ--AY-----ENNGDLJK 814

QY 651 -EQDPNKKVVVKT-----HHN-----SYALADGEHFI-----VDPDKLSS-----TIV 686

Db 815 AENAVEKIILKLTNQEDHYKLGIRFKLKKYEHSEIESPDQITKLDPKKHKALHNKGIALM 874

```

QY 687 DLKILKLEGLDYSTLROTTOTQFRDMMTAGKALLSKSNLRQGEKQKFLQBAQ----- 740
D 875 MLNKKKAIESPE-----KAIQIDKNYGTAYYQKGAEEKGDMQ 915
QY 741 ----PFLGRVDLDKAIKAKSKALVTKATQNGQLLERSIN-----KAVLAYNNSAIK 788
D 916 AFASFRNAYLNDKNPNYALKAGIVSNLGNFKQSEYLNFFNANAKPNEIAIYNLSIAK 975
QY 789 KANVKLELELDLLTGLVE-----GKGLAQATWQGVLLKPLPLP 831
D 976 PEN-NKLESLETINKAIDLNPEKSEYLYLKASINLNKKNYQVNSLYSLVIEKAP-ENT 1033
QY 832 EYVIGLVNVDYSGKLIYALD-MSDTIGEGQDAYGN-----PIL 870
D 1034 SAYINLAKAYEKGSKNSQAISTLEKIKNKNLALNGLYKKEKNYQKATEIFEKAI 1093
QY 871 NVDENEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAYHRAGIFQAIQN 928
D 1094 N--SDIEAKYNLATTLLIEINDNTRAKDLR-----EYTKLPNNPEALHALGIIENEN 1145

RESULT 49
Q9BJY0 PRELIMINARY; PRT; 2752 AA.
AC Q9BJY0;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE 235 kDa rhotry protein (Fragment).
GN Name=Illa.1;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21273138; PubMed=11378199; DOI=10.1016/S0166-6851(01)00253-5;
RA Khan S.M., Jarra W., Bayele H., Preiser P.R.;
RT "Distribution and characterisation of the 235 kDa rhotry multigene
RT family within the genomes of virulent and avirulent lines of
RT Plasmodium yoelii."
RL Mol. Biochem. Parasitol. 114:197-208 (2001).
DR EMBL; AF323442; AAK15625.1; -
DR InterPro; IPR006499; ReticulocyteBP.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
FT NON_TER
FT NON_TER
FT NON_TER
SQ SEQUENCE 2752 AA; 322608 MW; E834E3753168AF76 CRC64;

Query Match
Best Local Similarity 4.1%; Score 210; DB 2; Length 2752;
Matches 201; Conservative 162; Mismatches 355; Indels 262; Gaps 47;

QY 7 TVALTLTVSVVTH-NQEVFSVLKGPILKQTOASSI-----SCADYA 48
D 479 TISLILPXMDEPTNLNKMVQLNGKILKFKSISQIKNLNVSTVPERKEGFLSSIELA 538
QY 49 ESSGSKSLKI-----NETSGPVDVTVDLSDKRTTPEKIKMLAKGPRQELKAYTE 101
D 539 KSWKEKEEIIKKINEENETVQDIKIRELF-----KKYLDLAEEKYIEDLKLEIN 591
QY 102 N-----TESEKQITSGSQLESKES-----LSINKTVRSTSNWEICDRFTKNTLVGLSK 151
D 592 NKIKDITEKIKYKXKTVDLKKEVEKDVIYIDELAKQPP-----YQTEYEKKNTYIDTIK 647
QY 152 SGVEKLSQTDHLVLPSSQAAGDTQLIQVASFATPDCKTIAIABYTSRAGENGISQLVDVG 211
D 648 SDIKQIYVDIDLNYEMS-----SVLQENTINNENK--TELETIK 687
QY 212 KEIINEGEVFNYSLLKKVTIPTGYKHIGODAFVDNKNIAEVN--LPESLETISYAFAH 268
D 688 SKIDN---YNNK-----IKHLETEAVEKNNLNINETNNKNSLGSITLIEIKKIYGE 733

```

```

QY 269 LALKQID--LPDNLKAIAGELAFFDNQIT---GKLSLPQLM----- 304
D 734 ID-KELNKTLDNPKNEQELSKNINDYTKENDKLSVYKSKILLEIRYNYNDQINIDNTKGE 792
QY 305 RLAEAFPSN-HIKTIEFRGNSL-KVIGEASFQDNDLSQMLPDGLEKIE---SEAPT 357
D 793 EAKENYKSNHEHMTIPTNEYEISEAINEVVRMKNEL-----LSKVYKVPFDNIYN 844
QY 358 GNPQGDHYNRVLWTKSGKN---PSGLATENTYVNPDKSLWQESPE-----IDYTK 406
D 845 GNVDESH--NKFIDLTNKKNEVSDEKLSKHETKENGSKSLINETKKSIDIDYQINITLK 902
QY 407 WLEEDFTY---QKNSVTG-FSNKGLQKVRNKNLEIPKOHNGVTITEIGDNAFRNVDFON 462
D 903 KVDEYIKVCKRTKESITNFFSKQTTLKDLQNINTIKKIDPI-----EKPYKD-QPEN 955
QY 463 KTLKYVDLEEVKLPSTIRKIGAFAPQSNLKSFEASDDLE-----EIKEG 507
D 956 RLINKIS-----ELDKKPKDASLXDHESNNGLMEYFNNLKANLGNKKEKTLYHEFDEKEK 1011
QY 508 AFMNNRIETLELKOKLVT-----IGDAAPHINHIYALVLPESVQIEGRSAFRONGANNLI 562
D 1012 A-VNNIITIEDINKNISIEIYTSIYNIEEIEDAIGKNIKLLNDQVVKVKNVNTN 1070
QY 563 FMGSKVTLGEMAFPLSNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREAF 622
D 1071 LSGIKEKL-----KRYNFDLFGKEENIKYI-----NEIKINDINTLDQKXD 1112
QY 623 KKNHLKQLEVASALSHIAFNALDDNDGDEQFDNKVVVTHNSYALAGEHIVDPDKLS 682
D 1113 ----KSITLTIKTKTSESHIGEIKGTQDKLEK-----ADINTYXEDPKSIE 1156
QY 683 STIVDLEK-----ILKIEGLDYSTLROTTOTQFRDMMTAGKALLSKSNLRQGEKQK 735
D 1157 KKIENVVKKXKKNYKEINKL-----LNEISEIE-KDKTSLEE--LKNINLSYGRSL-- 1207
QY 736 LQEAQFFLGRVLDKAIKAEKALVTKATQNGQLLERSINKAVILAY--NNSAIKKNVYK 793
D 1208 --GNIFLEQIDEEK-----KKA-----ERTI-KAMEAYIEDLDNKKKS-D 1244
QY 794 RLEKELDLTLGLVEGKGLAQATWQGVLLKTPLPPEYVIGLVNVDYSGKLIYAL-- 851
D 1245 EIEKDMKIKMDINEEMKALNISNDNRNYHTKS-----KDHKKGISDIHDKSKSIIFNSK 1300
QY 852 -----DMSDTIGEGQK 862
D 1301 ESDINNIKNLEQENVSESRK 1320

RESULT 50
Q97K41 PRELIMINARY; PRT; 1227 AA.
ID Q97K41
AC Q97K41;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Uncharacterized protein, probably surface-located.
GN Ordered locus names=CAC1080;
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 824 / DSM 792 / VRM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RX DOI=10.1128/JB.183.16.4823-4838.2001;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucetce-Stamm L.A., Soucaille P.,
RA Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."

```







```

Db 441 ELGENKIFEDVLDGLIKLEVLDSNYIKDISLKNLTDIKELKLNKKNVSDISIVANN 500
QY 637 SHIAFNALDND-----GDEQFNKVVVVKTHNSYALADGE 672
Db 501 KNLQRLYINDNNTTKYLKDAKDLVLTANNKITSFEGELNLDLKEIH-----551
QY 673 HFIVDPDKLS-----STIVDLEKI-----LKLIEGLDYSTLRQTQTQTPRD-----713
Db 552 ---VDNNKISKDPLKNLKELETLSARTNVISDLKPIENLDYIKNLYLVENKISDISPLK 608
QY 714 -MTTACKALLSKNLR 728
Db 609 NMTGMLRLYLDKNNIK 624

```

## RESULT 54

```

RBPI_PLAVB
ID RBPI_PLAVB STANDARD; PRT; 2873 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN Name=RBPI;
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731; DOI=10.1016/0092-8674(92)90642-P;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites.";
RL Cell 69:1213-1226 (1992).
RN [2]
RP REVISIONS TO 1027-1036; 1110; 1278-1280; 1564; 1599-1607; 1804; 1861;
RP 1942 AND 2101-2103.
RA Tran T.M., Galinski M.R.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to
CC human reticulocyte cells.
CC -!- SUBUNIT: Homodimer (Potential).
CC -!- SUBCELLULAR LOCATION: Membrane-bound.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M88097; AAA29743.2; -.
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 22 Potential.
FT CHAIN 23 2873 Reticulocyte binding protein 1.
FT DOMAIN 23 2809 Extracellular.
FT TRANSMEM 2810 2830 Potential.
FT DOMAIN 2831 2873 Cytoplasmic.
FT SITE 2603 2605 Cell attachment site (Potential).
SQ SEQUENCE 2873 AA; 330524 MW; F4F0B9F2699E37B4 CRC64;

```

## Query Match

Best Local Similarity 4.1%; Score 208.5; DB 1; Length 2873;

Matches 214; Conservativity 198; Mismatches 398; Indels 377; Gaps 54;

```

QY 23 EVFSLVKPEILKQTOASSISGADYAESGSKLKNINETSQVDDTVTLFSDK-RTTPE 81
Db 228 QFNLEILELIPKDSQYHT-----YYEET--LKQVTEYTNLSKTLMDSCISEKQDMIL 279
QY 82 KITDNLAQ-----GPQRELKAVTE-----NTESEKQITSGSQLEQSKESLSINKTV 128

```

```

Db 280 EYEINYAKEKSIETETLGDKETKLSAVSQAYAKHLESYKGVLPKPLKINDIKNNAPSVLKDS 339
QY 129 PSTSNWEIC-DPITKGNITLVGLSKSGVEKLSQTDHLVLPQSAADGTQLIQVASFATPPDK 187
Db 340 YCKDN---CGEVQKNTNTRKNFPISSSDQYKMEAVYVIPSINDYTVLQKILS-----389
QY 188 KTAIAAYTSRAGENGESISOLDVDGKEIINE-----GEVFNYSYLLKKVTIPTGYKHIG 239
Db 390 -----ESNELGIDIOETVNSLKLGBEISEVSHLYVINSTLIDDAAKKL- 433
QY 240 QDAFVDNKNIAEVNIPESLETISDYAFAPHALAKQIDLPNLKAIAGELAFDNOITKLSL 299
Db 434 -ESINEEDESAEDLQK-----FEDNSKALANNYICFOYI-----467
QY 300 PROLMRLAERAPKSNHIKTIKPRGNSLKVIGEA--SFQDNDLSQMLPGLKIESEAF 357
Db 468 -KTLNEPIKAYESKVIKSNELLSTIIDLGKSAFALQESTFDQ-----ECCNKIATEA- 520
QY 358 GNPFGDDHYNRVVLTWTKSGKNPSGLATENTYVNPDKLWQESPE-----401
Db 521 -----EKKVDDAEDICEKN-----EQIYVEIPESEDETIIDDKINDLQDL 559
QY 402 IDYTKWLBEDFTYQKNSVTGFSNKGLOKVKRNKL-EIPKQNGVITITIG--DNAFRNV 458
Db 560 IDQKEYKDEIV---NNSEFISN-----RYKNIYENLKETYETELANDIGLENDTSKV 609
QY 459 DFQNTLRYDLEEVKLPSTIRKIGAF-----AFQSNLKSFEA 497
Db 610 NFYLMQIRKINTEKTKIDESLQTVKFKYKILDSKEKIYELKIEFKSVTEINRLQDGS 669
QY 498 SDDL--EIKEGAFMNNRIETLE-LKDKLVITGDAAFHINHIYAIYLPESVOEI-GRSAP 553
Db 670 ARDLHEEQIKE--ILDKMAKKVHLYKELLSLKSGSSVYPTENNELLNTASYDNMGFSAK 727
QY 554 RQGANNLIFMGSKVKTIGEMAFL-----SNLEHLDLSEQQLTE-----IP 596
Db 728 KEKADNDNALNSVYREDINALIBEVEKFVTENKESTLEMLKQEMEEKLODAKETPAK 787
QY 597 VQAFSDNALKEVLLPASLKTIRREAFKKN-----HLKOLEVASALSHIAFALDDN 647
Db 788 LNFVSDDKLTDVYTKMSAEVTVNAEGIKKIAQKQFENVHKKWKFSDAFS-TKFEALQNS 846
QY 648 -----DGD-----EQFDNKVVVK-----THNSYALADGEHF 674
Db 847 MQQYNQEGDAIEKHKQNRSEKEEYFKNESVEEDLSRETEEQYTKHKQNFRRKGE-- 904
QY 675 IVDPKLSTIVDLKILKILIEG-IDY-----STLR 704
Db 905 -----ISAEITNREVINIKIESQLNYGVIEKYSFLIGDQNEVSTAKALKEKIVSDSLR 958
QY 705 QTT---QTOFRDMTTA-----GKAL--LSKSN-----LRQEKQK 734
Db 959 DKIDQYEFKFKTSVAVENTVSTIOSLSKAIDSLKRLNGSINNCKYNTDIDLLRSKIKT 1018
QY 735 FLQEAQFFLGRVLDLKAIAKAEKALVTKK-APKNGQLLERSINKAVLAYNNSAIAKKNV 793
Db 1019 LREEVQKIEAETEGDKVVGENTTALLSLRDKMGKINEK-----LNDGLNSLDTK 1070
QY 794 RLEKELDLLTGLVEGKGPLAQATMVQGVYLLKTP-----PLPEYIIGLVNYPD-----842
Db 1071 ---KE-DLLKFYESKSKIHLSKQOKGP---QDPLNRIDEWEDIKREVDLNVNVQVISE 1123
QY 843 -----KSGKLIYALDMS---DTIGEG---OKDAVGNPI-----LNVEDNEGVHALAV 884
Db 1124 NKVTLFKNSVTVIEAMHSHINTVAHITSNNKNEILKSVKEVEDKLNLVQEONEDYK--KV 1181
QY 885 ATLADYEGLD-----IKTLNSKLQSLTSIRQVPTAAVHRA-----920
Db 1182 KNPENEKQLEAIRGSMKLEKVINKHVSEMTOLESTANTLKSNAKENEHLEBLNKTK 1241
QY 921 -----GIFOAIQNAAAEAFQLPKPCTHSEKSSSSSSANSKORGLQSN 963
Db 1242 GQMRDIYEKLIKIAEEL-----KEGTVELKDANEKANKVELEFERN 1283

```



```

RESULT 55
Q8RGSS
ID Q8RGSS5 PRELIMINARY; PRT; 197 AA.
AC Q8RGSS5;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE Hypothetical cytosolic protein.
GN OrderedLocusNames=FN0209;
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA DOI=10.1128/JB.184.7.2005-2018.2002;
RA Kapatal V., Anderson I., Ivanova N., Resnik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Busch G., Haselkorn R.,
RA Fongstein M., Kyrpides N.C., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL, AB010533; AAL94415.1; -.
KW Complete proteome.
SQ SEQUENCE 197 AA; 22356 MW; 460289F03260F5A4 CRC64;

Query Match 4.1%; Score 208; DB 2; Length 197;
Best Local Similarity 25.6%; Pred. No. 0.062;
Matches 61; Conservative 43; Mismatches 74; Indels 60; Gaps 4;

QY 402 IDYTWLEEDFYQKNSVTGFSNKGLOKVK--RNKNLSIPK-QHNGVITEIGDNAPRV 458
DB 1 MDONTWEYDDDFKGDDELKGMTQKQKVKVSGKTDLVPELTPDGLPKKIADNAF--- 57

QY 459 DQNKTLRKYLEEVKLPSTIRKIGAFQSNLKSFEASDDLEIKGAFNNRIETLE 518
DB 58 -----YRGLTSVIISTVESIGYDAFVCKLKEVKLPALNVIEGFAYRNKLTQVE 110

QY 519 LKDKLVITGDAAFHNIHAIYVLPESVQIEGSAFRQGNANLIPMGSKVKTGLGMAFLS 578
DB 111 FGSKVRLEPSPSFAMNELSELNLPETVEYIGASAFVK----- 147

QY 579 NREHLDLSEQQLTEIPQAFSDNALKEVLLPASLKTIREAFKKNHLKQLEVASAL 636
DB 148 -----NSLETVPSPKSVTKIDMYAFRNKNIHKVEVANSV 181

RESULT 56
Q64VN1
ID Q64VN1 PRELIMINARY; PRT; 520 AA.
AC Q64VN1;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DE Putative cell surface antigen.
GN ORFNames=BF1698;
OS Bacteroides fragilis.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YCH46;
RA Kuwahara T., Yamashita A., Hirakawa H., Nakayama H., Toh H., Okada N.,
RA Kuwahara S., Hattori M., Hayashi T., Ohnishi Y.;
RT "Genomic analysis of Bacteroides fragilis reveals extensive DNA
RT inversions regulating cell surface adaptation."

```

---

```

RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2004).
DR EMBL; AP006841; BAD48445.1; -.
SQ SEQUENCE 520 AA; 57220 MW; 17CF3336B968925F CRC64;

Query Match 4.1%; Score 208; DB 2; Length 520;
Best Local Similarity 21.5%; Pred. No. 0.24; Mismatches 203; Indels 148; Gaps 22;
Matches 121; Conservative 91;

QY 49 ESSGSKLKINETSGPVDVTDLFSKRTTPPEKIKONLAKGPREQELKAVTENSEKQ 108
DB 42 ELPGRYFIYFGEELDVSDMTLKVFSNSETSIVSKXDEITGPN-----NTVPEN-DQILE 96

QY 109 ITSGSQLEBQSKESLSLN-----KTPVSTSNWIEICDPITKNTLVGLSKSGVEKLSQTD 161
DB 97 VHKGSTVTFKIQVLINDIQAISIKTLPSKTVYTLGPELSLN----- 139

QY 162 HLVLPSQAADGTQLIQVASFAPTPDKKTAIAEYTSRAGEGEISQOLDVDGKEIINEGRV- 220
DB 140 -MWLEINVADGTIKENSAPSA-----DWQGFNSVPAQIQIVTLEDGKQVDFVOIL 192

QY 221 -----FNSYLLKKVTIPTGYKHHGQDAFVDNKNI--AEVNLPSLETIS 262
DB 193 PVKVDGDKVSVSDSDFTS-----ITPDDGIRTIGSKAF-ENKNIKASELLFPASLSTIE 246

QY 263 DYAFALHA-LKQIDLP-DNLKAIGELAFFNQITGKLSLPRQLMLRAERAPKS----- 313
DB 247 QAAPAYCRNLKIVDLSHTSIKELPEEAPLFSGIK-KIALPASLRIVGKEAFYGCCTDLNVI 305

QY 314 --NHKITEPRG-----NSLKVIGASP-QQNDLSQMLPDGLEXIESAP 356
DB 306 DISHTSVKELQNGAFGKSGISSISLPSTFKIVGTSAFIETKNLKELTILPESEVIDLEAF 365

QY 357 TGNPGDDHYNRRVVLTKSGKNPSGLATENTYVNPDKSLMQESPEIDYTWLEEDFYQK 416
DB 366 -----SGSSIQKVTLPNTIYHIDRSFY-NCPELITIE-----TYGT 400

QY 417 NSVTGFSNKGLOKVRKNKLEIPKQHGVTITEIGDNAPRVDPQNTLRKYDLEEVKLP 476
DB 401 RTTPS-----PVDRTAAIVSECFNHPK-----LTVLKIP 430

QY 477 STIRKIGAFQSNLKSFEASDDLEIKGAFNNRIETLEKDKLVITGDAAFHNIH 535
DB 431 ASIAKIGISALNKQVKTLILPVSVKALDFNAGNAVSLDEISLMSPTMTADYPPVPPR 490

QY 536 IYAIVLPESVQIEGSAFRQNGA 558
DB 491 IQKIRVPQNLVE---TVKQNK 509

RESULT 57
Q839L8
ID Q839L8 PRELIMINARY; PRT; 901 AA.
AC Q839L8;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DE Surface exclusion protein, putative.
GN OrderedLocusNames=EF0146;
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927; DOI=10.1126/science.1080613;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seeshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L.A., Brinkac L.M., Beanan M.J.,
RA Dougherty S.C., DeBoy R.T., Durkin S.A., Kolonay J.F., Madupu R.,
RA Nelson W.C., Vamathevan J.J., Tran B., Upton J., Hansen T., Shetty J.,
RA Khouri H.M., Utterback T.R., Radune D., Ketchum K.A., Dougherty B.A.,
RA Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant

```



657 LKEKQ-----AAQKABENTLSLKEVLN--LAKENLNQKQVAFKTTST---RSLSR 701

795 LEKELDLLTGLVBGKGLPQAQTWQGVYLLKTLPLPLPEYIGLNVTYFDRSGKLIYALDMS 854

702 LEN-----AQ-----PTYEKALN-----ELNKA 719

855 DTIGEGQKDAYGNPIILNVDBEDNGYHALAVATLADYEGDLIDKITLNSKLSQLSIRQVPT 914

720 EAAAVQAQEAVENTSLASLSELKE--QQA VATLTAYTQAQEDLSNAKLELQYQGVLRLELE 776

915 AAVHRAGIQAIQNAABAEOLLPKPGTHSEKSSSESANSDRGLO-----SNPKT 966

777 AQAEOQROEALQEQVAKSQQRLEBAKOSQTULVASATSDAKTPGGLQQLSFSKQKQPFKA 836

967 NRGHSA-----ILPRGTSKGS-----FVYIGLGYTSVALLSLITAIKKK 1006

837 QALTHSESRTKQVAKAPDPSLPHGTGEKNKWLAIAGLIFALLG--AAGIISFISRNEKK 893

RESULT 58

Q845L8 PRELIMINARY; PRT; 2167 AA.

ID Q845L8

AC Q845L8;

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Variable membrane protein precursor.

GN Name=vmp;

OS Mycoplasma hominis.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI\_TaxID=2098;

[1]

RP RP SEQUENCE FROM N.A.

RA Boesen T.;

RT "Gene and protein structure of the Mycoplasma hominis vaa adhesin.";

RL Thesis (2001), Department of Medical Microbiology and Immunology,

RL University of Aarhus, Aarhus, Denmark.

[2]

RP RP SEQUENCE FROM N.A.

RA Boesen T.;

RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ545046; CAD67982.1; -.

DR InterPro; IPR003900; KID\_repeat.

DR Pfam; PF02524; KID; 6.

KW Signal.

FT SIGNAL.

FT CHAIN

SEQUENCE 2167 AA; 244220 MW; C65BEC714CA00D58 CRC64;

Query Match 4.1%; Score 208; DB 2; Length 2167;

Best Local Similarity 20.6%; Pred. No. 1.7;

Matches 222; Conservative 159; Mismatches 401; Indels 294; Gaps 50;

QY 6 KTVALLPLTVSVVTHNQVFSL--VKEPILK-----QTQASSISGADYAESGSKSLKIN 59

DB 880 KKALIVQTLLKTKVDEKLITIEEKIKNALIKISNEETKLKSTESA-----KQQAQ 930

QY 60 ETSGPVDDTVTDLPSDKRTTPEKIKDNLAKGPQEQLKAVTENTSE-----K 107

DB 931 ET-----FNSKK-----QELKSLDFTTSEVDKAAEEKYKNT 963

QY 108 QITSGQLSQSKESISLNTKVPSTSNWEICDFITTKNTLVGLSKSGVEKLSDTHLVLP 167

DB 964 TIDNSTIKQIKQKTSDIENALSTLKEKTA--AKGDKAVELOKFNLVK--QLEELIAKE 1020

QY 168 QAAD--GTQLIQVASFAFTPKKT---AIAEYTSRAGE--NGEISQLDVGCKEINEGEVFN 222

DB 1021 DAKEVGTSKAEALKVNVADENSTLDAISKATKALEDVKAELTQNTDAKD--NATKTFN 1078

QY 223 SYL--LKKYTIPTGYKHIGQDAFVDNQNTAEVNLPESTLISDYAFAHLLAQIDLPDL 280

DB 1079 DKDELKLGLESSDAK-----IVDNKKGSDVLNNGSV-----DSN 1113



Qy	840	--YFDKSGKLIYALNMSDTIGEGQKDAYGNPILNVDENEGZH	880
		:    :      :      :	
Db	1831	QKISDKNKFLHALNET-----PIPNFTLKLBIYH	1860

**RESULT 60**

```

Q50B4
ID Q750B4 PRELIMINARY; PRT; 3645 AA.
AC Q750B4;
AD Q50B4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE AGR043Wp.
DE Q50B4;
GN OSFNames=AGR043W;
OS Ashbya gossypii (Yeast) (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX NCBI_TaxID=33169;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10895;
RA Lersch A., Brachat S., Voegelé S.E., Gaffney T., Philippsen P.,
RA Dietrich F.S.;
RL Submitted (SEP-2002; AAS54532.1; --
RL EMBL; AE016820; AAS54532.1; --
DR AGD; AGR043W; -.
DR AGD; AGR043W; PH.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00233; PH; 1.
SQ SEQUENCE 3645 AA; 409727 MW; 839C556DC514E73A CRC64;

```

Query Match 4.0%; Score 205.5; DB 2; Length 3645;

Best Local Similarity 20.0%; Pred. No. 4.6;  
Matches 239; Conservative 175; Mismatches 434; Indels 347; Gaps 54;

```

Qy 5 LKTVALTITVVTHNQEVPSIUKPELILKQATQSSSISGADYABSSGSKLKLINETS GP 64
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 494 LQSSGYKVIITISYEALSEKANLVASPPLEFLE-----EQAKNIGKVIIDOEYASL 545
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 65 VDDFTVTDLFSDKRTTPEKIKDNLAKGPREQELKAVTENTESEKOITSGSOLEQ--SKESLS 123
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 546 QDPSEVDIIISKAESHG---FVLGSKSAHQOMKLSIEQPTLEYLQ-----SKUQOHSMAIK 598
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 124 LNKTVPTSTNWEICDFITKGNLTVLGSKSGVEKLQSTDHLVLPQQAADGTOLIOVASFAP 183
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 599 LNDYALFENPTE--DYLSSKARL--LEKS-----LVNSELAEALRKISQEPSF-- 642
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 184 TPDKKTAIAEYTSRAGNGEISQLDVGKBIINEGEVFNVSLLKKVITPTGYKHI----- 238
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 643 -----DYLSSRAAANISCOLLPDQDFDALN-----KKANSPT--LEHLKGASA 682
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 239 GQDAFVDNKNIAEVLNPLESLETISDYA--FAHLALKQIDLDPNLKATGELAFPDNQITGK 296
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 683 ALDHVLVDKQDEYEHLLNPPLEKADLAATYKHALLSQHEHESLHPSAEL-----LTKR 736
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 297 LSLPQRLMLRAER-----AFKSNHIKTI-----EFRGNLSLKVI----- 329
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 737 LTBELDQVMLPEKKEYENLLSPTKHDI EQYAAACNHLYSDEDFKQLTSPSISVVIQASEL 796
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 330 GEASFQNDLSQLMLPDGLEKIESE----AFT-----GNPGDDHYNRNVVL 371
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 797 AHAVVPNEELEALKHPD--LETIKSHLSIAVTALPIEELGNKNKQISNPTLDHIS----- 850
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 372 WTAKSGKNPGLATENTVYNDPKSLWQES--PIDYTK-----WLEE--DFTYQ 415
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 851 -AKAQHQSVLLPIDTY---NKLLEDSDPSMDVMVKKHIEKKOKKELLSWLAQQLDHSIM 906
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 416 KNSVTGFSNGL-----QVKRNKNLBIPKOHNGVTITEIGDNAFRNVDPQNKTLR 466
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 907 KNSHAELLKRLDPSPTLEFLAEKAKAYDYLFEKKEHDIILSTWENFA---IDYLOTKAK 963
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 467 KYDLEEVKLPSTIRKIGAFAPQSNLKSFEASDDLBIKKEGAFWNNRIETLE----- 518
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

964	DB	HNHM	-----IVISADYYDELKQIVESPDGFLKKHASERN--LEILDSNVLLDLR 1011
519	QY	-----LKDKLVTIGDAAPHINHIYAI	-----VLPESVQEIQR 550
1012	DB	TVAESPSLDHLKNAAGAAAGYALISVSELEANTYANSPPLSHVSISSASTYKHEVIEATAL 1071	
551	QY	SAFRQNG	-----ANNLI-----FMGSKVYKTLGEMAF 577
1072	DB	NVLRKADAPSLDLHLQSCAAAI DYMLVMAEYNKLVLTLEAPPIEFLEEKARAHQOQSLIP 1131	
578	QY	SNRLHLDLSEKQOLITPIVQAFSDNA--LKEVLLPASLTKTIREEAFKQOHLKOLEVAS-- 634	
1132	DB	ENDL-----IALTSKLEBPPELLEYISQKARGLKHVLLPL-----EBHKSLLLENANN 1176	
635	QY	ALSHIAFNALDDNDGDEQFD--NKVVVVKT-----HHNSYALADGSHFTVD----- 677	
1177	DB	PSLEHLTKAEDNGYVVVPYDEHKRIIVLSEDDPTLEHLSKAKARG--HVVVNESRYEDLQ 1235	
678	QY	PD-----KLSSTIV-----DLEKILKUIEGLDYSTLRQTTQTOTQFRDMTWTAGKA 720	
1236	DB	SLANSPDLNHLCDKAKLGGCTIVNNSQDALLKSIIEQPSLYLKEKAEAYAHNCVPTKEY 1295	
721	QY	LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAERKALVTKAT--KNGOLLERSINKAV 779	
1296	DB	L-----EKKLVEDPDLAYLEKKATEARETLIKUDAYKAL 1330	
780	QY	LAYNNSAIKANVRKLEKELDLTLGLVEGK--GFLAQATMVQGVYLLKTPLPUPYYIG 836	
1331	DB	LEKADS---PSLQHLSDKAESLDHLVVARNEFNDLRLSAAQHSDDLAKLEAPSLEYL-- 1384	
837	QY	LVNVFDGSKGLIYALDMSDTIGEQCKDAYGNPILNVDBEDNEGYHALAVATLADYEGLDIK 896	
1385	DB	-----KEKTVGYELVPLDITYSD--LKQTFDPDPSLDLFLNKSKEKGFDPOLVAITEYKEL--LR 1436	
897	QY	TILMSKLSQLTSLIRQVTAAYHRAGIEQAQTQNAAEAEQLLPKPGTHSEKSSSESANSK 956	
1437	DB	RAVDPITDEKADV-----LRKTRVILSQDTADKLQKTAESP-----SKSFLAEKAAATM 1485	
957	QY	DRGLQSNPKTNRGHRSAILPRTGSKGSFVYGIILGYTSVAL-----LSLI--TAIKK 1005	
1486	DB	GLLL-----VEKEKYVDVLLKSSDKDTILSSIQGFGFIPPIPELSLLKNSAVDK 1535	
RESULT 61			
Q8CP76	ID	Q8CP76	PRELIMINARY; PRT; 9439 AA.
AC	DT	08CP76;	
DT	01-MAR-2003	(TREMBlrel. 23, Created)	
DT	01-MAR-2003	(TREMBlrel. 23, Last sequence update)	
DT	01-OCT-2003	(TREMBlrel. 25, Last annotation update)	
DE	EbAa	protein.	
GN	OrderedLocusNames	=SE1128;	
OS	Staphylococcus epidermidis.		
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.		
OX	NCBI_TaxID	=1282;	
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 12228;		
RX	PubMed=12950922;		
RA	Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,		
RA	Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,		
RA	Xuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.,		
RT	Genome-based analysis of virulence genes in a non-biofilm-forming		
RT	Staphylococcus epidermidis strain (ATCC 12228)."		
RL	Mol. Microbiol. 49:1577-1593(2003).		
DR	EMBL; A016747; AA004725.1; .		
DR	HSSP; P02976; 1HOT.		
DR	InterPro; IPR011439; DUF1542.		
DR	InterPro; IPR011490; FIVAR.		
DR	InterPro; IPR002988; GA.		
DR	Pfam; PF07564; DUF1542; 4.		
DR	Pfam; PF07554; FIVAR; 62.		

RESIST 61

Q8CP76	PRELIMINARY;	PRT; 9439 AA.
ID	Q8CP76	
AC	Q8CP76;	
DT	01-MAR-2003 (TREMBLrel. 23, Created)	
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)	
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)	
DE	Ebha protein.	
GN	OrderedLocusNames=SEI1128;	
OS	Staphylococcus epidermidis.	
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.	
OX	NCBI_taxID=1282;	
[1]		
RN	SEQUENCE FROM N.A.	
RP	STRAIN=ATCC 12228;	
RC	PubMed=12950922;	
EX	Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J., Ran Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z., Qin Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.; "Genome-based analysis of virulence genes in a non-biofilm-forming Staphylococcus epidermidis strain (ATCC 12228)."; Mol. Microbiol. 49:1577-1593(2003).	
DR	EMBL; AE016747; AAC04725.1; --	
DR	HSSP; P02976; 1HOT.	
DR	InterPro; IPR011439; DUF1542.	
DR	InterPro; IPR011490; FIVAR.	
DR	InterPro; IPRO02988; GA.	
DR	Pfam; PF07564; DUF1542; 4.	
DR	Pfam; PF07554; FIVAR; 62.	



```
Db 809 K--KDAESGVLEGAFFKLNKNETGVVGETTKTNKGVVVFENLVPGKYTLLEETKAPCY 866
QY 480 RKIGAFAPFNNLKSPEASDDLEEIKEGAFMNRIB-----TLEKDKLVITIGDAAFHI 533
Db 867 KAVEV-----TVEVNVANVVKQFVMEKVTGQFEIIVKVDANDAKVLSDAEFTV 917
QY 534 NHIYAVLPESVQIEGRSFRQNGANNLIIFMGSVKVLTGEMAF-----LSNRLRH 583
Db 918 YKDGKKVAELKTDESCK-----VMSPKL-PLGEYTVKTKAPGYKLSNKEWK 964
QY 584 LDLSEKQLTETPVQAFSDNALKEVLLPASLKTIRREAFKQKHLKOLEVASLASHIAFNA 643
Db 965 VTIQNEKEV--VKVEAENERIL-----GSLQIITKD--DKQAKLSLGAEP-----T 1007
QY 644 LDNDGDE-----QFQNVVVVTHNSYALADGEHPIVD-----PDKLSST-----IVDL 689
Db 1008 LKDAQNVVKEGITTDKSIGVVD-----GLVGEYTLLEETKAPGEYELTKQVHVTVDGE 1063
QY 690 KILKIEGLDYSTLRTOTTQFRDMTTAGKALLSKSNLRQGBKQKPLQBAQFFLG-----744
Db 1064 KI-----VDVKVTSKSLGQFEIVKVD-----DKAKVLSDAEPEVYKDGKK 1106
QY 745 -----RVD-LDKAIA-KAEKALVTKATKNGQ-----LLERSINKAVLAYNNSAIKKANVKRL 795
Db 1107 VETLRDTKGTGVISQKLEPGTYTLKETKAPQGYKLLKEBIEVVVEANKVQVQVENAKEL 1166
QY 796 -----EKELDLTLGLVEGK-----GPLAQATWVQGVLLKTPLP 829
Db 1167 GSLQVKKDAESGVLEGAFFKLNKNETGVVGETTKTNKGVVVFENLVPGKYTLLE-ETK 1225
QY 830 LPSEYIIGLVNYPD-----KSGKLIYALDMSDT-----IGRG 860
Db 1226 APEGYKALEVTVNVVNVANTVVKQVLEKVEKIKGVQVEITKVDAITDKKLAVPEI 1285
QY 861 QKDAVGNPILNVDENEGHALAVATLADY-----EGLDIK-----TILNSKLS-- 904
Db 1286 LKD--GTKIDTLTDTKNGKATSKLEPGDYILKEVQAPGEGYELSDKGIEFTISNOKIEV 1343
QY 905 --QLTSIRQVPTAAVHRAGIFQA-----IQNAAEAE-----QLLPKPGTSEK-- 946
Db 1344 KLIQITNEKTSKGPENPGGETETPGGETETPGGETETPGGETETPGGETETPGGETETPG 1403
QY 947 -----SSSESANSKDRGLQ 961
Db 1404 EBTGKPGEBETKPGEBETKPGEBETKPGEBETKPGEBETKPGEBETKPGEBETKPGEB 1463
QY 962 SNPKTNRGRHSAILPRTGSKGSFVYGLGYTSVAL 996
Db 1464 TLPERKQOGASHAQLPATCHDMNVL-PFIGFALVLL 1497

RESULT 63
Q99QR6
ID Q99QR6 PRELIMINARY; PRT; 2481 AA.
AC Q99QR6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE FmtB protein.
GN Name=fntB#mrp; OrderedLocusNames=SAV2160;
OS Staphylococcus aureus (strain Mu50 / ATCC 700699).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699;
RX MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;
RA Kuroda M., Ohba T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
```

```
RA Kanahisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus";
RL Lancet 357:1225-1240(2001).
DR PIR; D90011; BAB58322.1; -.
DR EMBL; AP003364; BAB58322.1; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR000890; Acetate_kin.
DR InterPro; IPR011439; DUF1542.
DR InterPro; IPR011490; FIVAR.
DR InterPro; IPR005877; Gpos_Ysirk.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF07564; DUF1542; 18.
DR Pfam; PF07554; FIVAR; 1.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
DR PROSITE; PS01075; ACETATE_KINASE_1; UNKOWN_1.
KW Complete proteome.
SQ SEQUENCE 2481 AA; 263765 MW; E1EAB99B81665E8 CRC64;

Query Match 4.0%; Score 204.5; DB 2; Length 2481;
Best Local Similarity 18.5%; Pred. No. 3;
Matches 206; Conservative 159; Mismatches 403; Indels 343; Gaps 45;

QY 10 LTLTIVSVTHNOEVPSLVKPEILKQ-----TQASSSI-----SGADYAESGSKLKI 58
Db 1293 LQATQTVINNDQATNEEKEAAIQQLATAVTDAKNNTAATDDGVTADKAGNSIQS 1352
QY 59 NETSGPVDDTVTLPSDKRTTPEKIKDLAKGPREQELKAVTENTSEKQITSGSOLEQS 118
Db 1353 TOPATAVKSNAKNEVDQAVTQONQAIQNTT-----GATTEKVAADLVKKEKAY 1404
QY 119 KESLSLNTK--VPSTSNWEICDF--ITKGNLTVLGSKSGVEKLSQDHLVLPSSQAADGTQ 174
Db 1405 QDILNAQTNDVTQIKDQAVADIQGITADTTIKDVAK-----DELATKAN----- 1449
QY 175 LIQVASFAPTPDKKTAIAEYTSRAGENCEISQLDVGDKKEINEGEVFNFSYLLKKVTIPTG 234
Db 1450 -----EQKALIAQTADATTEKEQANQVDAQ-----LTQG 1480
QY 235 YKHIGQDAFVNKNTAEVNLPSLETSIDYAFAPHLAQLAQIDLPDLNKAIGELAPFDNOIT 294
Db 1481 NQNTENASQIDVNTAKDNALQAIQPIQ-----ASTDVKTNARA-ELTENVKQKIT 1530
QY 295 GKLSLPLQMLRAERAPFNSNHIKTIEPRGNSLKVIGEASFQNDLSQLMLPDGLEKIBSE 354
Db 1531 EIL-----NNNETTNEEGNDIGPV-RAAYE-----EGLNNINRA 1564
QY 355 AFTGNPGDDHNNRVVLTGSKGNPSGLATENTYVNPDKSLWQESPEIDYTKWLEBEDFTY 414
Db 1565 TTTGDD-----VTTAKDTAVQVQQLHANPVKPKPAGK-----KELDQAAAD 1604
QY 415 QKNSVTGFSNKGLOKVKENKLEIPKQHNGVTITIEGDNAFNVDFQNKTLRKYLEEVK 474
Db 1605 KKTQIEQTPNASQQEINDAKQ-EVDTELN-----QAKTNVD-QSSTNEYVD----- 1648
QY 475 LPSTIRKIGAFQFQSNLKSFEASDDLBEIKEGAFMNNRIETLE--LKDKLVTIGDAAFH 532
Db 1649 -----NAVKEGRKAKINAVKTFSEYKDKALAKIEDA--- 1678
QY 533 INHIYAVLPESVQIEGRSFRQNGANNLIIFMGSVKVLTGEMAFISNRLRH-----DL 586
Db 1679 -----YNAKVNEADN-----SNASTSSEIAEAKOKLAELKQTDADQNV 1715
QY 587 SEQKQLTEIPVQAFSD-NALKEVLLPASLK-----TIREAPFK 624
Db 1716 NQATSKDDIEVQIHNDLNDINDIYIPTGKGSATTDLYAYADQKQKNTSADTNATQDBKQ 1775
QY 625 NHLKQLE--VASALSHIAFNALDNDGDG-----QFQNVVVVTHNSYALA- 669
```

```

Db 1776 QAIKQVQDNVQTALENIN-NGVDNGDVGDDALTOGKAAIDTIQVDATVVKPANKAQAIEAKAE 1834
Qy 670 DGEHFIIVDPDKLSS--TIVDLKILKLEGLDYSLRTQTQTPFRDMMTAGKALLSKSNL 727
Db 1835 DTKESIDHSQDTAEKTEALAMIKQITDQAKQGITDATTAEVEKAKAQGLEAFDNIQI 1894
Qy 728 RQGEKQKFLQEAQFFLGR---VLDKKAIAKAEKALVT-----KKATKQQLLERSIN 776
Db 1895 DSTEKQKAELEETALDQIEAGVNVVDADATTEKEAFNALEDILSKATED--ISDQTN 1952
Qy 777 KAVLAYNNSAIKANVVKRL-----EKELDLTLGLVEGKGPLAQATMVQ 820
Db 1953 AEIATVKNLSALEQLKAGRINPVVKNKALEAIREVVNKQIEII-----KNADADASAKE- 2005
Qy 821 VYLLKTLPLPEYIYGLNVYFDK-SGKL-----IYAL-----DMS 854
Db 2006 --IART-----DLGRYDFRFDKLDKTQNTNEVAELQNVITPAIEAIVPQNDPDAN 2054
Qy 855 DT-IGEGOKDAYGNPILNVDENEGVALAVATLADYEGLDIKTILNSKLSQTSIROVP 913
Db 2055 DTNNGTDNDATANSANATPENTQPNVSETT--DNKADASPPTPNNSDAATGETTVT 2112
Qy 914 TAAVHRAGIFQAIQNAAAEAEQLLP-----KPGTHSEKSSSES 952
Db 2113 SATDADKQKQANNSADASTNSPTMDNDVTSKPEVESTNGTDDKPVTTEDNATPAES 2172
Qy 953 A---NSKQGLQSNPKTNGRHSAILPRTGS 980
Db 2173 TTNNSTTTATNENAPTQ---STATAPTAS 2200

RESULT 64
Q7A4B1
ID Q7A4B1 PRELIMINARY; PRT; 2481 AA.
AC Q7A4B1;
DT 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE FmbB protein;
GN Name=fmbB(mrp); OrderedLocustNames=SA1964;
OS Staphylococcus aureus (strain N315);
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003136; BAB43253.1; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR000890; Acetate kin.
DR InterPro; IPR011439; DUF1542.
DR InterPro; IPR011490; FIVAR.
DR InterPro; IPR005877; GpoS_YSRK.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF07554; DUF1542; 18.
DR Pfam; PF07554; FIVAR; 1.
DR Pfam; PF04650; YSRK_signal; 1.
DR TIGRfam; TIGR01167; LPTXG_anchor; 1.
DR TIGRfam; TIGR01168; YSRK_signal; 1.
DR PROSITE; PS01075; ACETATE_KINASE_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 2481 AA; 263765 MW; E1EAB99B81665E8 CRC64;

```

```

Query Match 4.0%; Score 204.5; DB 2; Length 2481;
Best Local Similarity 18.5%; Pred. No. 3;
Matches 206; Conservative 159; Mismatches 403; Indels 343; Gaps 45;

Qy 10 LTLTVSVVTHNQVFSVLVEPILKQ-----TOASSSI-----SGADYAESSQSKLKI 58
Db 1293 LQATQTVINNDQATNEEKEAAIQQLATAVTDAKNIIATATDDNGVDVDTAKDCKNSIQS 1352
Qy 59 NETSGPVDVTDLPFSKRTTPEKIKONLAKGPQEQLKAVTENTESEKQITSSQLEQ 118
Db 1353 TQPATVAKNAKNEVDQAVTTQNAIDNTT-----GATTEKNAAKDLVLAKEKAY 1404
Qy 119 KESISLAKT--VPSTSNWEICDP--ITKGNLTGLVSGVGEKLSQTHLPLPSAADGTQ 174
Db 1405 QDIUNAQTNDVTQIKQAVADIQGIATDTTIKQVAK-----DELATKAN----- 1449
Qy 175 LIQVASFAPTKKTAIAEYTSRAGEISQLDVGKEIINEGEVFNYSLLKKVTPTG 234
Db 1450 -----EQKALIAQTADATTEKEQANQVDAQ-----LTQ 1480
Qy 235 YKHIGQAFVQDNKIAEVLNPLESLETISDYAFALHALKQIDLPNLKAKIGELAFDQIT 294
Db 1481 NQNIENAQSIDDVNTAKDNATQAIIDPIQ-----ASTDVKTNABA-ELLTEMONKIT 1530
Qy 295 GKLSLPQLMELAEAFKSNHIKTIKFRGNSLKVIGRASFOQNDLSQLMLPDGLEKTESE 354
Db 1531 EIL-----NNETTNEKGNIDIGPV-RAAYE-----EGLNNINAA 1564
Qy 355 AFTQNGDDHYNNRVLWTKSGKNPSGLATENTYVNPDKSLWQESPRIDYTKMLEEDFTY 414
Db 1565 TTTGD-----VTTAKDTAVQKVCQLHANPVKPKAGK-----KELDQAAAD 1604
Qy 415 QKNSVTGFSKGLQKVKRNKLEIPKQHNQVVTITEIGDNAPRVNDFQNKTLRKVDLEEVK 474
Db 1605 KKTQIEQTPNASQOEINDAKQ-EVDTBLN-----QAKTNYD-QSSTNEYVD----- 1648
Qy 475 LPSTIRKIGAFAPQSNLKSPEASDLEEKEGAFMNNRIETLR--LKDKLVTTIGDAAFH 532
Db 1649 -----NAVKEGKAKINAVKTFSEYKDKALAKIEDA--- 1678
Qy 533 INHIYAVLPESVQIEGRSAPRQNGANNLIPMGSKVKTGLGEMAFLSNRLHL-----DL 586
Db 1679 -----YNKAVNEADN-----SNASTSSSIAEAKQKLELKQTDQNV 1715
Qy 587 SEQQLTEIPVQAFSD-NALKEVLLPASLK-----TIREBAPKK 624
Db 1716 NQATSKDDIEVQIHNDLDNINDYTIPTKCKESATTDLYAVADQKKNISADTNATQDEKQ 1775
Qy 625 NHLKQLE--VASALSHIAFNALDNDGDGE-----QFDNKVVVTKHNSYALA- 669
Db 1776 QAIKQVQDNVQTALENIN-NGVDNGDVGDDALTOGKAAIDTIQVDATVVKPANKAQAIEAKAE 1834
Qy 670 DGEHFIIVDPDKLSS--TIVDLKILKLEGLDYSLRTQTQTPFRDMMTAGKALLSKSNL 727
Db 1835 DTKESIDHSQDTAEKTEALAMIKQITDQAKQGITDATTAEVEKAKAQGLEAFDNIQI 1894
Qy 728 RQGEKQKFLQEAQFFLGR---VLDKKAIAKAEKALVT-----KKATKQQLLERSIN 776
Db 1895 DSTEKQKAELEETALDQIEAGVNVVDADATTEKEAFNALEDILSKATED--ISDQTN 1952
Qy 777 KAVLAYNNSAIKANVVKRL-----EKELDLTLGLVEGKGPLAQATMVQ 820
Db 1953 AEIATVKNLSALEQLKAGRINPVVKNKALEAIREVVNKQIEII-----KNADADASAKE- 2005
Qy 821 VYLLKTLPLPEYIYGLNVYFDK-SGKL-----IYAL-----DMS 854
Db 2006 --IART-----DLGRYDFRFDKLDKTQNTNEVAELQNVITPAIEAIVPQNDPDAN 2054
Qy 855 DT-IGEGOKDAYGNPILNVDENEGVALAVATLADYEGLDIKTILNSKLSQTSIROVP 913
Db 2055 DTNNGTDNDATANSANATPENTQPNVSETT--DNKADASPPTPNNSDAATGETTVT 2112
Qy 914 TAAVHRAGIFQAIQNAAAEAEQLLP-----KPGTHSEKSSSES 952

```



```

Db 2113 SATDDAKDPQANNSSADASTNPTMDNDVTSKPEVESTNNGTDDKPVTTEDNATPAES 2172
QY 953 A--NSKDRGLQSNPKTNRGRHSAILPRTGS 980
Db 2173 TTNNSTTTATNENAPTG---STATAPTTAS 2200

RESULT 65
Q9ZHL0
ID Q9ZHL0 PRELIMINARY; PRT; 4919 AA.
AC Q9ZHL0; Q7BY44;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)
DE Large supernatant protein 2.
GN Name=lgpA2; OrderedLocusNames=HD1156;
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=35000;
RX MEDLINE=99030326; PubMed=9811662;
RA Ward C.K., Lumley S.R., Latimer J.L., Cope L.D., Hansen E.J.;
RT "Haemophilus ducreyi secretes a filamentous hemagglutinin-like
RL protein."
RL J. Bacteriol. 180:6013-6022 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=35000HP / ATCC 700724;
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT "The complete genome sequence of Haemophilus ducreyi."
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF057696; AAC9761.1; -.
DR EMBL; AE017154; AAP96012.1; -.
DR PIR; T31105; T31105.
DR GO; GO:0004197; P:cyteine-type endopeptidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008619; Fil_haemagg.
DR InterPro; IPR008638; Haemagg_act.N.
DR InterPro; IPR003951; Peptidase_C58.
DR InterPro; IPR006473; Peptidase_C58_yt.
DR Pfam; PF05594; Fil_haemagg; 10.
DR Pfam; PF05860; Haemagg_act; 1.
DR Pfam; PF03543; Peptidase_C58; 1.
DR TIGRFAMs; TIGR01901; adhes_NPXG; 1.
DR TIGRFAMs; TIGR01586; yopt_cys_prot; 1.
KW Complete proteome.
SQ SEQUENCE 4919 AA; 542602 MW; 5779201455CA69A0 CRC64;

Query Match 4.0%; Score 204.5; DB 2; Length 4919;
Best Local Similarity 20.0%; Pred. No. 7.8;
Matches 250; Conservative 181; Mismatches 415; Indels 405; Gaps 65;

QY 10 LTLTVSVVTHNQEVPSLVKEPIILKQTQA-----SSISGADYAESGSKLK----- 57
Db 598 VTLSKSTLSAGELTFKPKYKNTLNDSELAANNLSLASHNVTLNKSLSAQAADIKIA 657
QY 58 -----INETSQPVDDTVTDLFSPDKRTTPEKI-----KNL----- 87
Db 658 VNLTNDTTELTAKNLDINSTITNNGIAGIFAN--ITTEKLNKKEKALILAEQNLNPT 715
QY 88 -----AKGREQELKAVNTESKQITSGSQLESLSLN----- 125
Db 716 VNGSHVENKGDIVSKDQATVTFESKNSDFTSNGSKLVNAQNQLKVNANNFTISQGDITLI 775
QY 126 -----KTVPTSNWEICDPTTKNTLVG-----LSKSGVB-----KLSQT 160
Db 776 GNVTLNASGTFNTSGNLTTVKTLDVGDIQNFTNKGNTLVGDLHKTSKITNDGKLISI 835

```

RESULT 66  
Q7RTL4  
ID Q7RTL4  
AC Q7RTL4;

PRELIMINARY; PRT; 17903 AA.











```
Db 6023 NNAMTKBIEQAKAQLA--QALQDIKDLVKAKEDAKNAIKAL-----AN 6064
Qy 931 ABAEQLLPKPGTHSEKSSSS--ESANSKDRGLQS--NPKT-----NRG 969
Db 6065 AKRDQINSNPDLTPQKAKALKEIDEAEKRALQNVENAQTIDQLNRG 6111

RESULT 71
Q99U54 PRELIMINARY; PRT; 6713 AA.
AC Q99U54;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Ebha protein.
GN Name=ebha; OrderedLocusNames=SAI267;
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N315;
RX MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;
RA Kuroda M., Ohca T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kihara S., Goto S., Yabuzaki J.,
RA Kanahisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus".
RL Lancet 357:1225-1240 (2001).
DR EMBL; AP003133; BAB42527.1; -.
DR PIR; B89921; B89921.
DR InterPro; IPR011439; DUF1542.
DR InterPro; IPR011490; FIVAR.
DR InterPro; IPR02988; GA.
DR InterPro; IPR006530; YD.
DR Pfam; PF07564; DUF1542; 8.
DR Pfam; PF07554; FIVAR; 44.
DR Pfam; PF01468; GA; 46.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 1.
KW Complete proteome.
SQ SEQUENCE 6713 AA; 722339 MW; AF6BEDE226BE4888 CRC64;

Query Match 4.0%; Score 203; DB 2; Length 6713;
Best Local Similarity 19.5%; Pred. No. 14;
Matches 232; Conservative 183; Mismatches 396; Indels 376; Gaps 64;

Qy 34 KQTQASSISGAD-----YABSSGSKLKINETS GP-VDDTVTLFSDKRTTPBKIKON 86
Db 5050 QQTETAGSFNEDPKQDAYQAQVAKDLINQTNPTLDKSVQBLTQAVTT---AKDN 5106
Qy 87 L--AKGPREQELKAVTEN-----TESEKQITSGSQLEQSK 120
Db 5107 LHGDQKLARDQQAQVTTVNALPNLHNAQQUTDAINAAPTETVEAQQVQATLHDHAME 5166
Qy 121 SLSLNKTVPSTNWBICDIPITKGNLTVLGSLSGVEKLSQTDHLVPSQAADGTQLIQVAS 180
Db 5167 TLK-NKV-----DQVNTDKAQPNTYESTDKKEAVD-----QALQAAQ 5203
Qy 181 FAFTPD-----KXTAIAEYTSRAGENGESQLVDVGKEIINEGEVFNLSVLLKKVTIPTGY 235
Db 5204 SITDPTGNSNANKDAVEQALTKLQS--KVNEL--NGNERVAEAKTQAKQTIDQLT----- 5254
Qy 236 KHIGODAFVDNKN-----IAEV-----NLPSLETISDYAFALHAKQ-IDL--- 276
Db 5255 -HLNADQIATAKONTDQATKQPIAEVLVDQATQLNQSMDQLQQAQVNEHANVEQTIDYTQA 5313
Qy 277 -PDNLKAIGE-LAPFDN---QITGKLSLPRQLMRL--AERAF-----KSNHIKTIE 320
Db 5314 DSDKQKAYKQATADAENVLKNANKQVQDQALQNLINAKQALNGDERVALAKTNGKHDID 5373
```

```
Qy 321 PRGSLKVI GEAS P-----QNDLSQLM-----LPDGLKEKIESAFTCNPGDDHYN 367
Db 5374 -QLNALANNAQQDGPGRIDQSDNLNQIQIVDEAKALNRAMDQL-SQEITGNR----- 5426
Qy 368 RVVLWTKSGKNPSGLATENTYVNPDKSLWQSPS-IDYTKW-----LE 409
Db 5427 -----TKGSTN-----YVNADTVQKVYDAVDKAKQALDKSSQNLTARQVIKLN 5472
Qy 410 EDFTYQKNSVTGFS-----NKGLOKVKR 432
Db 5473 DAVTAALKALNGEBRLNRRKAEALQRLDQLTHLNAQRLAIQQINNAETLNKASRAINR 5532
Qy 433 NKMLE-----IPQHNGV-----TITEIGDNAPRV-----DFQNKTLRK 467
Db 5533 ATKLDNAMGAVQVYIDEQHLGVISSTNYINADNLKANYDNAINAANAHELDKVGQNAIAK 5592
Qy 468 YDLEEVK-----LPSTIRKIGAFQSGNNL-----KSFEASDDLEE 503
Db 5593 ABAEQLLKQNIIDAQNALNGDQNLANAKKANAFVNSLNLNQOQDPLAHKAINNADTVSD 5652
Qy 504 IKEGAFNNRIETLELKDCLVTIGDAAAPHINHIYIVLPESVQBIGRSAFRQNGANNLIF 563
Db 5653 VTD--IVNQID-----LNDAMETLKLVDNEIPNAEQTVNYQNAADDNAKTNPDD 5700
Qy 564 MGSVKVTL-----GEMAFLSNRLHLDLSEQKQ-LTEIPVQAFSD---NALKE 607
Db 5701 AKRLANTLLNSDNTNVNDINGAIQVNDAIHNLNGDQRLQDAKDKATQISNOALANKLKE 5760
Qy 608 VLLPASLTREEAFKKNHLKQLEVASALSHIAFNALDNDGDQFD---NKVVVKTHN 664
Db 5761 I--EASNATDQDKLIANKKAEEL--ANSIINI--NKATSNQAVSQVQTAGNHATEQVHAN 5815
Qy 665 SYALADGEHFIVDPDKLSSTIVDLK-ILKLEIGLDYS-----TLRQTTQTQ 710
Db 5816 EIPKAK-----IDANK-----DVDKQVQALIDEIDRNPNTDKKQALKDRINQILQOG 5864
Qy 711 FRDMTTAGKALLSKSNLRQGEKQ--KFLQEAQFL-GRVDLDKAIKAKALVTK----- 762
Db 5865 HNDINNA-----LTKEBIEQAKAQAQALQDIKOLVKAKEDAKQDVQKQVALIDEIQNP 5920
Qy 763 -KATNGQOLLERSINKAVLAYNSAIKKANVKRLKELDLTLGLVEGKPLAQATMVQGV 821
Db 5921 NLTDKEKQALKDRINQ-ILQCHNGINNAMTK---EBIE-----QAKAQLAQA----- 5964
Qy 822 YLLKTPPLPBYITGLNVYFDKSK-LIYALDMSDTTIGEGQKDAYGNPILNVDED--NEG 878
Db 5965 --LKEIKDLVKAKENAKQDVQKQVQALIDEIQNPNTDKKQALKDRINQILQCHNDI 6022
Qy 879 YHAL-----AVATLADYEG-LDIKTILNSKLSQTSIROVPTAAHYHRAGIFQAIQNA 930
Db 6023 NNAMTKBIEQAKAQLA--QALQDIKDLVKAKEDAKNAIKAL-----AN 6064
Qy 931 ABAEQLLPKPGTHSEKSSSS--ESANSKDRGLQS--NPKT-----NRG 969
Db 6065 AKRDQINSNPDLTPQKAKALKEIDEAEKRALQNVENAQTIDQLNRG 6111

RESULT 72
Q723X5 PRELIMINARY; PRT; 1775 AA.
AC Q723X5;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Cell wall surface anchor family protein.
GN OrderedLocusNames=LMOF2365_0350;
OS Listeria monocytogenes (serotype 4b / strain F2365).
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=265669;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15115801; DOI=10.1093/nar/gkh562;
```





```
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF01463; LRRC1; 1.
DR Pfam; PF00560; LR_1; 33.
DR PRINTS; PR00019; LEURICHRPT.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
FT NON_TER 1 1316
FT NON_TER 1316 1316
SQ SEQUENCE 1316 AA; 145688 MW; C64B0D7172AF6336 CRC64;

Query Match
Best Local Similarity 4.0%; Score 201.5; DB 2; Length 1316;
Matches 240; Conservative 185; Mismatches 452; Indels 401; Gaps 59;

QY 26 SLVKEPILKQTOASSI-----SCADYAESGSKLKINETSQVDDT----- 68
DB 3 SILNAPVRKDFSNALSIGTKALFGVAEYLTELRLANNLLGDSLPIFSTDVLQALQ 62

QY 69 ---VTDLFDKRTTPKIKDNIAKPRE-----DELKAVTENTSEKQITSGSOLEQSK 119
DB 63 SLKVLDSGNRIIA---LEESIFDGNRLVELYLQNKIATVPVAAVKELSLKLL--SL 117

QY 120 ESLSNKTVPSFSNW---EICDFITKGNLVLGSKGVKLSQTDHLVL-----PS 167
DB 118 RGNRIESLPDAFPANKLERLDL--RNNRIRSLKSKAFANLASMKEVLLAGNQLSHIDE 175

QY 168 QAADGTOLIOVASPAFTPKTAIAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLK 227
DB 176 RALAGMDVLQKLDLS-----DNLLSEFPSEA--LGSVVSLKVLNLSLNNIGKLESNHLAQ 228

QY 228 KYTIPTGYKHIGQDAFVNKNIAEYNLP-----ESLETISDYAFALHAL 271
DB 229 MKSL-----QIIDISERTIATI-LPOTFREQLLLKYVLSLNSLTIEDDAFEGH-- 277

QY 272 KOIDLPDNLKAI---GELAFPDNOITGKL-----SLPQLMLRAEPKSNHKTIE--- 320
DB 278 -----DNLQTLIRDNINILFTPGSALGRLPKLSNLYLDFNRVA--ALSSSLKSIOPEN 329

QY 321 -----FR-----GNSLKVTGEASFPQND----- 338
DB 330 IYLSLSRNVIRELPDGPFAFRKLIYLDISGSLGVIGEDTFKLEGTLLEIKLSFNRI 389

QY 339 ---LSQLMLPDGLEKTESEAFNCPGDD-----HYNNRVVLWTKGKNPSGLATENTVYN 390
DB 390 ASLRKFVLP-----KLRLDLSANSIDDLAIDSFHSLHLLLYLNMGNHVGQVT--RTMIY 444

QY 391 PKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGKQVK-----RNKNLEIP-KOHNG 444
DB 445 P-----LTKLQVIDLSHCGKMLQADLPFNNTDLRIALLNHNQ 482

QY 445 VTIIRIG-----DNAP-----RNVDFQNTLRKYDLBEVK--LPSTI 479
DB 483 LRVIEGTFLNLLNLFDSLAGNLEQLRPSFVNVTNLRTLNLRQNALQELRADMFTE 542

QY 480 KRIGAFAPQSNLKSFEASDDLE---ETKEGAFMNRIT-----LELKD--- 521
DB 543 TALEVLDASDNALKGF-ATNLKIIPRLKVLANNRLEVFPPELIADLHLEVIDLAGN 601

QY 522 KLVITGDAAF-----HINHIYAIVLPSVQBIQSRAPRONGANNLI-FMGSVKYTLGEMAF 576
DB 602 RUTTIQOLDPGRLANLRELY--LRANTIDSVDMAFHNSSQLQILDAAANRLERSRSF 659

QY 577 LSN-REHLIDLSEOK-----QLEIIPVQAFSD----- 602
DB 660 EGTLLDRDLSDNKLTLGLPEQVLAKSRVQLRSLVSLAGNKLASFVPAFQGHDTLQTL 719

QY 603 -----NALKEV-----LPLASL-----KTIRBEAPKKNHLKOLE 631
DB 720 DLSRNAIREVPRAASHMLMINAKHVDPSYNYPLTPEAIATVLGPKVTRTLNLAGTVREL 779

QY 632 VASA-----LSHIAFNALDNDGDQFDKNVVK---THNSYALADGEHFIVDPDKL 681
DB 780 MLETPYLSLNSLNHNSIAV---SAKAPEKVTLLERLDLSHNAIGDADG----- 825
```

```
QY 682 SSTIVDEKILKLEGLDYST--LRQTTQTQFRDMTTAGKALLSK-SNLROGEKQKP--- 735
DB 826 ---LRDIWPKLALGYLDLSSNPRTITASTFERLDGLTELYRELPELTRLEKNAPKL 882

QY 736 ---LQEAQFF-LGRVDLDKAIKAEKALVTKATKNGQLLERSINKA-----VLAYN 783
DB 883 KDLTILQAYQPAKGYIDVQGIQVQELPALAAVDAVDEAKDANLESQOLQVLLHHPKLTLLGLH 942

QY 784 NSAIKKNVKKLEKELDLTLGLVEGKGLQAQATWVGYYLLKTP-----LPLP-BYYIGLN 838
DB 943 GYALQ-----SLSGAPAGLRNKVLTVALHNTSLTALPPALLPLPRSAHIDFS 991

QY 839 VYFQSKGLI--YALDMSDTTIGEGKDAYGNPILNVDEDNEGYHALAVAT-LADYEGLDI 895
DB 992 IAGSKVGTGLTPQFLLSSLDLDRKNSIQIDGLATNPVRCDCQAPAFRRWILATKVQDVRCASP 1051

QY 896 KTIILNSKLSQ-----TSIROVPTAAHYRAGIFQAIONAA 930
DB 1052 PTVAGKLTVEGDSLVDCGRSKATSTQPAASSTGGSTRTPTISFYNYTSEVQMTNAA 1111

QY 931 ABAEQLLPKPOTHSEKSSSESANSKDRGLQSNPKTNRGHSAILP-----RTGSKGSFV 985
DB 1112 ATTSEDI-----IWSMPVPVSSSTTTRTKGSKSVS-----SAAMPPLKOMPSSANDDTLI 1161

QY 986 YGILGYTSVALLSLITAI 1003
DB 1162 IGIVGVVVVFIFLLIICI 1179

RESULT 74
Q869B8
ID Q869B8 PRELIMINARY; PRT; 1885 AA.
AC Q869B8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Kinesin-related protein K4.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98359834; PubMed=9693369;
RA de Hostos E.L., McCaffrey G., Sugang R., Pierce D.W., Vale R.D.;
RT "A developmentally regulated kinesin-related motor proteins from
RT Dictyostelium discoideum."
RL Mol. Biol. Cell 9:2093-2106(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Suyama E., Sutoh K.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB102780; BAC56912.1; -.
DR HSSP; P33173; 1161.
DR DictyBase; DDB0191404; kmd.
DR GO; GO:0005875; C:microtubule associated complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
DR ATP-binding; Microtubule; Motor protein.
SQ SEQUENCE 1885 AA; 218215 MW; 15686AED3B007EC7 CRC64;

Query Match 4.0%; Score 201.5; DB 2; Length 1885;
Best Local Similarity 20.4%; Pred. No. 2.8;
Matches 213; Conservative 169; Mismatches 338; Indels 323; Gaps 54;

QY 21 NOEVFSL-----VKPEILKQTQASS-----SISGADYAESGSKLKINETSQP 64
DB 626 NOELLEIDQSKQSIQLNDKLLKLETKQSKQSIQLNLQIDIESSSKNKSF----- 679
```



```
QY 65 VDTVTDLSDKRTTPKIKDNLAKGPREQELKAVTENTSEKQI-----TSGSQLEQS-K 119
Db 680 --ENVLGVPEKSVRLAERLED-----KYFTKEIESKQIETLANSYLQLETTVQ 726
QY 120 ESLSLN-----KTVPTSTN-----W-----EICDFTKGNLTVG 148
Db 727 OQLNLNQSQQKQIQSLNNDIEQFKLVVPLKQDVNGYFOENQMFQKQIIELEBKNTLTD 786
QY 149 LSKSGVEKLSQTDHLVLPQADGTOLIQVAFPTPKKTAIAEYTSRAGENGESOLD 208
Db 787 LQKE-VEQNYLN--TLEQORNDQYQI-----EINQLTTEY-----NNQIQLE 827
QY 209 VDGKEIINEGEVFNFSYLLKKVT-----IPTGYKHGQ-----DAP-VDN 246
Db 828 STNQKL--QTQLYN--LLANATQSTQTLQSQLOTSQKEIDTLTNEIEQLKNQYDIIRVDN 883
QY 247 KNIAEVLNPESLETISDYAFALHALKQIDLPNLKAIGELAFDPNQITKLSLPQLMRL 306
Db 884 DNLSK-----ESLE-----LQIILSKTQLEQLSLAQOQ-KGNIEIIOQL--- 924
QY 307 ARAFAKSNHKTIEFRGNLSKLVIGEASFQNDLSQLMLPDGLEKIESEAFNGPGDDHYN 366
Db 925 -ESIIVDNO-QSID-----QKIEFQSQODN-----OSIKQS-----YN 957
QY 367 NRVLWTKSGKPNPSGLATEN---TYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGF 422
Db 958 QLESTLTLAQSBNRLTENKQFITSLEIKSLFNSIQOQKQETIQLEFEFQEKQFDSL 1017
QY 423 SNKGQKVRKNRKNLEIPKQHGVTITEIGDNAPRVDFONKTKRYDLEEVLPSTIRKI 482
Db 1018 LTNYNQLFKYNDLATSNEISNLEFQPKKDS-----NQSIQSLESLSLKS----- 1065
QY 483 GAFAPQSNLKSFEASDDLEEIEKEGAFWNRLETLE--LKDKLVTIGDAFHNHYAIV 540
Db 1066 -----ENDNL-----LQSSLLKSQLESIEKQKODQLIPI-----QLE 1098
QY 541 LPESVQIEGR--SAFRQNGANNLIFMGSKVKTIGEMAPLSNLEHLDLSEQKLEIPVQ 598
Db 1099 LESKCKELSKLSQSEBQTKQ---VTQLLISVDQVKISTNKLSEQISDRNEINNLK 1154
QY 599 AFSQNALKEVLLPASLK-----TIREAFKQKHLKQLEVASALSHPALND 645
Db 1155 AIEINALKEENI--SLKQOLTKLKKAPKQTDREKDMIKKELEKLE-----KFNAID 1205
QY 646 DNDGDEQFQNVKVVKTHNSYALADGEHEIIVDPKL-SSTIVDLKILKL-----LEGD 699
Db 1206 AK-LKQAIQDKQTIQSEKQSL-----EREIKDKRSHTSTETELDKLKKTHLAADVKSD 1259
QY 700 YSTLRQT-----TOTQPRDMTTAGKALLSKNLRQGEKQKQLEQAPFLGRVLDLKA 753
Db 1260 FIALNKSVEILTKSQEQLKSTIEBESDLSKNIELEKQEL-----VTLNQ--D 1308
QY 754 KAEKALVTKKATNGQLLERSINKAVLAYNNSAIKAN---VKRLEKE-LDLTLGLVEG 808
Db 1309 KLEK-----EKKTNQLES DSHSATIKLENVENQITOLTSEIIDLKSKFQEB 1354
QY 809 KGFLAQATWQGVYLLKT-----PLPIPEYIGLVNVPDKSGKL--IYALDMSDT 856
Db 1355 KSSESNTIKQEBINLKESNDLNQOLTNDNLKLTQKLSDLKVFEDKSKQLMSTRSESNDT 1414
QY 857 IGEQGDQAYGNPTLVNDENEGHALAVATLADYEGLDIKT-----ILNSKLQSLSIRQV 912
Db 1415 IKELQES-----IISKDKERQ-LTSEQVLKIDQ--INLKTWEYNDLNSQCQOLT----- 1461
QY 913 PTAAYHRAGIFQAIQNAABEQ 935
Db 1462 -----KTLQNVKSSNEQ 1473
PRT; 1578 AA.
RESULT 75
Q9AV25 PRELIMINARY;
ID Q9AV25
```



Db 1902 AIGQIDQDRSNAQVDTASINLQTIHDL-----DVHPIKKPDAEKTINDILARVLTALVQ- 1955  
 QY 698 LDYSTLRQTQTQFRDMTGTAGKALLSKSNLRQGEKQKFLQEAQFGLGRVDLQKALAKAEK 757  
 Db 1956 -----NYRKVSRENK-----ADALK 1970  
 QY 758 ALVTKATKNGQLLERSIN---KAVLAYNNSAIKANVRLEKELDLLGLVEGKPLAQ 814  
 Db 1971 AITALKLQMDLELKTARTNADVAVLRFNVALSDIEAIVTEKENSLLR--IDN---IAQ 2025  
 QY 815 ATWQGVLLKTPPLPEYVIGLVYFDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDE 874  
 Db 2026 QT-----YAKFAIATPEQLAKVKVLIDQ-----YVAD-----GNRM--IDE 2060  
 QY 875 DNEGHALAVATLADYEGDIKILNSKLSQTSIRQVPTAAYHRAGIFQALQNAAAEAE 934  
 Db 2061 D-----ATLN-----DIKHQTFIVDEILAIK-LPAEA-----TKVSPKEI 2095  
 QY 935 QLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG 982  
 Db 2096 QPAPKVCTPIKKEETHESKRKEVE-----LNTGSEG 2127

RESULT 77  
 Q7RNW8  
 ID Q7RNW8 PRELIMINARY; PRT; 2911 AA.  
 AC Q7RNW8;  
 DT 01-MAR-2004 (TREMBLrel. 26, Created)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Acetyl-CoA carboxylase 1 precursor-related.  
 GN Name=PY01695;  
 OS Plasmodium Yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxId=73239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17XNL;  
 RX PubMed=1236865; DOI=10.1038/nature01099;  
 RA Carlton J.M., Angluoli S.V., Suh B.B., Koolij T.W., Perteau M.,  
 Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
 Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,  
 Cho J.K., Quackenbush J., Sedegah M., Shoabli A., Cummings L.M.,  
 Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
 Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
 van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
 Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 Carucci D.J.;  
 RT "Genome sequence and comparative analysis of the model rodent malaria  
 parasite Plasmodium yoelii yoelii.";  
 RL Nature 419:512-519(2002).  
 CC -! CAUTION: the sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL: AABL01000457; EAA21055.1; --  
 DR HSSP; P24182; 1BNC.  
 DR GO; GO:0003343; C:biotin carboxylase complex; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004075; F:biotin carboxylase activity; IEA.  
 DR GO; GO:0016874; F:ligase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR005482; Biotin carb C.  
 DR InterPro; IPR000089; Biotin lipoyl.  
 DR InterPro; IPR000022; Carboxyl trans.  
 DR InterPro; IPR005481; Phase L.N.  
 DR InterPro; IPR005479; Cphb synth L\_D2.  
 DR Pfam; PF02785; Biotin carb C.1.  
 DR Pfam; PF00364; Biotin lipoyl; 1.  
 DR Pfam; PF01039; Carboxyl trans; 1.  
 DR Pfam; PF00289; Cpsase L chain; 1.  
 DR Pfam; PF02786; Cpsase L\_D2; 1.  
 DR PROSITE; PS00866; CPSASE\_1; 1.

DR PROSITE; PS00867; CPSASE 2; UNKNOWN 1.  
 SQ SEQUENCE 2911 AA; 337275 MW; F2A9083552BD9E02 CRC64;  
 Query Match 4.0%; Score 201; DB 2; Length 2911;  
 Best Local Similarity 19.0%; Pred. No. 5.4; Mismatches 364; Indels 286; Gaps 44;  
 Matches 192; Conservative 167;  
 QY 1 MKKHLKTLVALTIT-TVSVV--THNQEVFSLVKEPIKLTQOASSS----- 41  
 Db 1382 VKKYLTLNLSITNTINKMSTTPNEELINDISENNLSITNTMHNDSIVNSGNSDSILSVSR 1441  
 QY 42 -----ISGADYAESGSKLKLINETSQVDDVTDLFSDKRTTPPKI----- 83  
 Db 1442 TNNISEEKIDNNKLLISYDNKYNQSKIVSSEIDFFVNNT-----KDKKLDKMKIISD 1496  
 QY 84 -----KDLAKGPREQELKAVTENTESEKQITSSQLEQSKESLSLNKTVSTSNWEIC 137  
 Db 1497 TEMSRSDNDSOSSNR-----TNTNNEKIELNFNDKCKE-----NEFFLELSNMKRI 1546  
 QY 138 DPITKGNLTAVGLSKSGVEKLSQTDHLVLPSSQAADGTQLIQVASFAFTPDKTAIAEYTSR 197  
 Db 1547 EYLLKGYX-----QDYKCFED-----FLKDXNNFDKNLVI 1578  
 QY 198 AGENGEISQLDVGDKEIINEGEVFNLSYLLKVTIPTGYKHIGQDAFVNNKNIAEVLNPS 257  
 Db 1579 LDKFTYNNHL-FSKKEIISEVEIY-YIL-----YKSIS-----DKK-----KQ 1614  
 QY 258 LETTSDYAFALHALKQIDLDPN-----LKAIGELAFPDNQITGK-LSLPRQ 302  
 Db 1615 YEIIHAYTNYLSIKLILNLSKILNLSKQTNLSLQILNKIKILKEFGKIFGSIILLRY 1674  
 QY 303 LMRLAERAFKSNHIKTIEFRGNS-----LKVIGEASFQDN-DLSQMLPDGL 348  
 Db 1675 IILYEQNEKIHYES-SFQYNKQMEDMKQSLQNYIRFLTSSWMDYKIDQILNNTNY 1733  
 QY 349 EKIESEAFTPGDDHNNRVVLMTSGKNPSGLATENTYV-----NPDKS 394  
 Db 1734 EE-KIKFLKSPSDIHYSLPFLFKT---ENVYQSMFNTYLNHLYKYSTIKAKHSPNCL 1789  
 QY 395 LW-----QESPEIDYTKLEEDFTYK-----NSVTGF 422  
 Db 1790 VFSINNHYNNVMLNLEDITIEQPKSMLDILYEMKSDKINTFHIINKNETFFNSLQNY 1849  
 QY 423 SNKGLQKVKR-----NKLEIPKHNGVTITIGIGNAFNVDVFQNKTLRKYLEEVK 474  
 Db 1850 LDNAFKKVTYLYINYFNKGYGEIHK-----INNAYKEREKIQDNLEKSIETNK 1900  
 QY 475 LPSTI-RKIGAFQSNLKSFEASDDLE-----EIKEGAFMNNRIETLEKD 521  
 Db 1901 NDSTINNKIYLLPFEEIILTSNTASNFELKNVQGVFNKTKILGTYKNNKYTSLPAQ 1960  
 QY 522 KLVITGDAAFHNIHYAIVLPESVQEGISAFRQGANLIPMGSKVTLGEMAFSLNRL 581  
 Db 1961 RIIDINDLQKQVDSNINIKYDEQ-ENVGMT---KNGKNNILL--SKNIENGQM---NNQI 2011  
 QY 592 EHLDLSEOKLTE-----IPVQAFSDNALKEVLLPASLKTIRE-----BAFKK- 624  
 Db 2012 YDIPMSLKESEDIESIYRLNPSRDIKISSNIYHIVLPNN-SNIEQINGIKIYKNF 2070  
 QY 625 -NHLKQLEVASALSHIAFNALDDNDGDFQFNKVVVVKTHNSYALADGEHFIVDP-DKLS 682  
 Db 2071 LNKYNELFIENYNNVYIKVYKEIEEKNNETKLELQKLLHLFDNEKEVIEKVDKIN 2130  
 QY 683 STIVDLKILKLEGLDYSTLRQTTQTOFRMTAGKALLSKNSURQGEKQKFLQEAQFF 742  
 Db 2131 PVIID-ALYLRKRGARDVDTI-----YAYDFVKLINIALNRINNTLNDNFYINISVKEF 2183  
 QY 743 LGRVDLDK-----AIKAEKALVTKKATNGQLLERSINKAVLAYNNSAIKKNV 792  
 Db 2184 --KIDYSRENSSEINDTVSGVKEKESQPNKNDKQSKWLFENFDNL-----SDE 2230  
 QY 793 KRLEKELDLALGLVEGKGLAQATWVGQVYLLKTPPLPEYVIGLVNLYF 841

```
Db 2231 IRIKSLYLSNSLEIGQKMSVIGLLMNV---KT-----DEYKEGREDIAP 2272

RESULT 78
ID Q7YWF0 PRELIMINARY; PRT; 3080 AA.
AC Q7YWF0
DT 01-OCT-2003 (TremBrel. 25, Created)
DT 01-OCT-2003 (TremBrel. 25, Last sequence update)
DT 01-MAR-2004 (TremBrel. 26, Last annotation update)
DE Normocyte binding protein 2a.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_taxid=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVO;
RX MEDLINE=20402589; PubMed=10920203; DOI=10.1073/pnas.160469097;
RA Rayner J.C., Galinski M.R., Ingravallo P., Barnwell J.W.;
RT "Two Plasmodium falciparum genes express merozoite proteins that are
related to Plasmodium vivax and Plasmodium yoelii adhesive proteins
involved in host cell selection and invasion.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:9648-9653(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVO;
RA Rayner J.C., Barnwell J.W., Galinski M.R., Tran T., Huber C.S.,
RA Corredor V.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY138499; AAN39446.1; -
SQ SEQUENCE 3080 AA, 363854 MW, 73PBD7DC5648968 CRC64;

Query Match 4.0%; Score 201; DB 2; Length 3080;
Best Local Similarity 19.7%; Pred. No. 5.8;
Matches 231; Conservative 189; Mismatches 395; Indels 356; Gaps 59;

Qy 32 ILKQTQASSISGADYAESGSKLNETS-----GPVDDTVTLFSDKKTTEKIKDN 86
Db 2025 LVKDCKELEISDALY-----DKIQTTSVINRENDISNNI-DIVSNKLNLEIDAIQYN 2076
Qy 87 LAKGPREQELKAVTENTSEKQI--TSGSQLEQSKS-----SLSLNTKTPSTSNWEICDFIT 141
Db 2077 FEK-----YKEIFDVEYKTLDDTKNAYIVKKAELKNVDINKTKE-----DLDI 2122
Qy 142 KGNLTVLGSKGVEKSLDTHLVLPQAAADGTQLIQVAFPTDPKTAIAEYTSRAGEN 201
Db 2123 YFNLDLELEK-----LTLSSNEMETKIVQ-----NSYNSFSFIN 2158
Qy 202 GEISQLDYDGK-----EINNEGEVFN-----SYLLKKVTIPTG----- 234
Db 2159 KNINDIDKEMKTLIPMLDELLNEGHNDISLYNFIRNIQIKIGNDKINREQENDTWIC 2218
Qy 235 -----YKHIGQDAFVNK-----NIAEYNLPESLETISDYAPAHALKQI 274
Db 2219 FEYIQNNYFNKSDISIFNKYDDHDKVDNYISNNIDVKNKSL--LSEH-----VINAT 2271
Qy 275 DLPDN-LKAIGELAFDNOITGKLSLR--QLMFLAERAPK-----SNHKTIEFR--- 322
Db 2272 NIENIMTSIVEI-----NEDTEMNLEETQDKLELYENFKKKNIIINNKKIVHFNKLG 2327
Qy 323 --GRSLKVIGEASFOVDLSQLMPLDGLKTESEAFNGDDHYNRRVLTGSKGKPS 380
Db 2328 EENSLETYSISTFNFKINETQNDILK-----NEFNN---IKTKINDKVK 2371
Qy 381 GLATENTVVPDKSLQWQSPET-----DYTKWLEEDFTYQKNSVTGFSNKGLOKVKR-- 432
Db 2372 EL-----VHVDSTLTLESIQTFNNLYGDLMSNIQDVVYKE-----DINNVELKKVKLYI 2420
Qy 433 -----NKNLEIPKQHG--VTITEIGDAPRNVDPQNK 463
Db 2421 ENITNLLGRINTFIKELDKYQDENNGIDKYEINKENNSYIILKKEKANNLKENF---SK 2477
Qy 464 TLKRYDLEEVKLPSTIRKIGAPAFQSNLKSFEASDDLEEKEGAFMNNRIE-----TLEL 519
```

```
Db 2478 LLQNIKNETEL-----YNINNIK-----DDI--MNTGKSVNNIKQKESNLPL 2519
Qy 520 KDKLVITIGDAAFHINHI-----YAVLPESVQEIQRSAFRQNGANNLIPMGSK 567
Db 2520 KEKLFQMEMLLNININNETKRISNTDVTYNI---TLQDIENNKKEN--NNM-----N 2569
Qy 568 VKTUGEMAFLSNRLEHLDLSEQKQLTEI-----PVQAFSDN---ALKEV----- 608
Db 2570 IETIDKL-----IDHIKIHNEKIQABILIIDDAKRKVKETITONKAFNEITENYNNEN 2623
Qy 609 --LAPASLKTIREEAFKKNLKO--LEVASALSHPAFNALDDNDGDFQFNKVVVKTHHN 664
Db 2624 NGVKSANVIDEATYLNNELDKFLKLNELLSH-----NNDIKDLGKELKLEEBE 2677
Qy 665 SYALADGEHFIVDPKLSSTIVOLEKILKIEGLDYTLRQTTQTQFRDMMTAGKALLSK 724
Db 2678 R-----KERERLEKAKQEBERKERERTEK--EKQEKERLEREKQELKK-----EALKKQ 2725
Qy 725 SNLRQGEKQK-----FLOBAQFFLGRVLDKAIKAEKALVTKKATKQGL-----LERSI 775
Db 2726 EQERQEQQKEEALKRQEQE-----RLQKEBELKQEQERLERE--KQSLQKQEBELKQ 2779
Qy 776 NKAVLAYNNSAIKANKVRLKELDL-----LTGLVEGKGPLAQATMVQGVVLL 824
Db 2780 QERL--QKEEALKRQEQERLQKEBELKQEQERLERKKEIHLAEREQHIKSKLESMDVKII 2837
Qy 825 KTLPLPEYIYGLNVYFDKSGKLIYALDMS-----DTIGEQQKDAYGNPILNVDEDEG 879
Db 2838 KDELTKEDKELIKN-----KDKLHRSLEQKWLKHLQNLISLKIDSLNKNDEVIKDNETQ 2893
Qy 880 HALAVAT-----LADYEG-----LDIKILNSKLSQTS--IRQVPTA 915
Db 2894 LKTNILNSLNQYLNLRKELNEIIEKEEYENQKKILHNSQNLVNDLSQKTNRLVDIKPT- 2952
Qy 916 AYHBAGIFQALQAAAABQLL-----PKPGTHSEKSSSSSSANSKDRGLQSNPK--TNRGR 970
Db 2953 ---KHGDIYTNKSDNTEMLITSKEKDETESTKRSQTDHTNSESSTDDTNDTNRNFSR 3009
Qy 971 HSAILPRTGSGSFVYIGLYTSVALLSLIT 1001
Db 3010 SKNLSVAIYTAGSVALCVLIFSSIGLLIKT 3040

RESULT 79
ID Q6GGX3 PRELIMINARY; PRT; 10746 AA.
AC Q6GGX3;
DT 05-JUL-2004 (TremBrel. 27, Created)
DT 05-JUL-2004 (TremBrel. 27, Last sequence update)
DT 05-JUL-2004 (TremBrel. 27, Last annotation update)
DE Very large surface anchored protein.
GN Name=ebh; OrderedLocusNames=SAR1447;
OS Staphylococcus aureus (strain MRSA252).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_taxid=282458;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15213324; DOI=10.1073/pnas.0402521101;
RA Holden M.T.G., Feil E.J., Lindsay J.A., Peacock S.J., Day N.P.J.,
RA Enright M.C., Foster T.J.J., Moore C.E., Hurst L., Atkin R., Barron A.,
RA Bason N., Bentley S.D., Chillingworth C., Chillingworth T.,
RA Churcher C., Clark L., Cotton C., Cronin A., Doggett J., Dowd L.,
RA Feltwell T., Hance Z., Harris B., Hauser H., Holroyd S., Jagels K.,
RA James K.D., Lennard N.A., Line A., Mayes R., Moule S., Mungall K.,
RA Ormond D., Quail M.A., Rabinowitsch E., Rutherford K.M., Sanders M.,
RA Sharp S., Simmonds M., Stevens K., Whitehead S., Barrell B.G.,
RA Spratt B.G., Parkhill J.;
RT "Complete genomes of two clinical Staphylococcus aureus strains:
evidence for the rapid evolution of virulence and drug resistance.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9786-9791(2004).
DR EMBL; BX571856; CAG40444.1; -
GO GO:0016020; C:membrane; IEA.
```

[illegible]

```
QY 282 AIGELAFD-----NQITGKLSLPROLMRLAERAFKSNH--IKTIEFRGNSL 326
Db 783 KI-----FDEINQKYDDLTQNVQNDTKIVLRQ-----ENLKGKYEIKKENIQDRDI 833
QY 327 KYVIGASFQDNDLSQLPDGLEKIESAFTPGDGHYNNRVVLWLTGSKGNPSGLATEN 386
Db 834 KYIGE-----NNDGNNNNNDNII--NGNNHTG----- 859
QY 387 TVNPDKSLMQSPRIDYT-----KWLEEDTYQNSV-TGFSNKGLOKVRNKNLEIP 439
Db 860-----DYNVTLFHYTDLWNSQFTRTKENIQTNINN--IKSLIRNLQNEL- 902
QY 440 KOHNGVTITEIGDNAP-----RNVDFONKTLRKYDLEEV-----KLPSTIR 480
Db 903--NNTNTLKSNGSHFYDKIHTLEELKILTQEDINDKNVRKQYDIENTYQNDLHTEELIK 960
QY 481 KIGAFAPSNLKSPEASDLEEKEGAFMNNRIETLEKOKLVIGDAAFHNIHYAI- 539
Db 961 NITSYIKINILNIIIS-----IKQTYNNKNSISLKLKINNLT-NSTQEVINQIKIP 1014
QY 540--VLPSVQIEGRS-----AFRONGANNL-----IPMGSKVKTGLGEMAPL 577
Db 1015 TNLLEPHIKQKSLSELNITFMQYDKLNEVINNLTYKSKDSLOFYINEKNKN-----I 1068
QY 578 SNRLEH-----LDLSEKQOLTEIPVQAFSDNALKEVLLPA-----SLKTIREEAPK--- 623
Db 1069 NNNNDHNVNDYNDIKEN-----QIYKNKLYECIQTKKOIDELNYDQLFKNIS 1120
QY 624-----KNHL-----KQLEVASALSHI----- 639
Db 1121 QNNYNNHLSLFIHSLNNHMSLIPQVTKYGHKNQOILSDINDVIKQNEHKBSTYNLDTNSI 1180
QY 640-----AFNALDNN--DGBQEPDNKVVVTHHNSVALAGEHFIVDP----- 678
Db 1181 QLIKEIQYFLQIPHILQENITTFENQYKDLIIKNNHKINNKLKDIITHIVINDYTLQEQ 1240
QY 679-----DKLASTIVDL-----EKILKLLEGLD-----Y 700
Db 1241 NNINYELQNKIKQIKRVNEVFTNNYNNQOILNYSQAQDNFPNIPMKFQNNINDINRKRY 1300
QY 701 STLRQTTQ-----TOFRDMTTAGKALLSKNLROQEKQKFLQEAQFFLGRVDL 748
Db 1301 NVQKITEILINSYDIINYNKNNIKDMYQQLKNIQOQLNTTETQLNHKIQNNHF----- 1354
QY 749 DXAIAKASKALVTKATKNGQLLERSINKAVLAYNNSAIKANVRKLELDLGLVLEG 808
Db 1355--KYFYKSNQTSISIVKNIQNEKLKIQEFNKKIQHPK-----BQTQIMINKLIQ- 1400
QY 809 KGPLAQATWQGVYLLKTPPLPEYVIGLVYFDKSGKLIYALDMSDTIGEGQKD---AY 865
Db 1401-----PSNIHLHKMLPITQOQANTHLRNEHTK-----NDTKSYNMKDEENAM 1444
QY 866 GNPILNVDEBDEGYHALAVATLADYGL-----DIKTIILSKLS 904
Db 1445 GYGTTN--ERKNSGTNDMNSTIGDNTVLTNDQBERGKDTSRNNNIHTDEKINNHETN 1503
QY 905 QUTSTRQVPTAAHYRAGIPQAIQNAABAEQLL-----PKPGTHSEKS 947
Db 1504 EOYPKHEPNPNYHQKDEKISLQHTKINTTSQRTIDSSNMDRNNRYNTSSQKNNLHTNN 1563
QY 948 SSSSESANSKDRGLQSNPKTNRGHSAILPRTGSKGSP--VYGILGYTSVALLSLTA 1002
Db 1564 SNSRYNNNHDK--QNEHKYNOQKSS-----GKNSYWRIFYAGGITAVLLICSSTA 1612
```

RESULT 81

Q6BU09

ID Q6BU09 PRELIMINARY, PRT; 2042 AA.

AC Q6BU09, 25-OCT-2004 (TReMBLrel. 28, Created)

DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)

DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)

DE Similar to sp|25386 Saccharomyces cerevisiae YDL058w US01.

```
GN ORFNames=DEHA0C09658g;
OS Debaryomyces hansenii CBS767.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=284592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG Genolevures;
RA DuJon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisarame A., Boyer J., Catolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi R.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest J.M., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozler-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT *Genome evolution in yeasts.*;
RL Nature 430:35-44 (2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG Genoscope;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382135; CAG86127.1; --
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008565; P:protein transporter activity; IEA.
DR GO; GO:0006886; P:intracellular protein transport; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000357; HEAT.
DR InterPro; IPR003900; KID_repeat.
DR InterPro; IPR006955; Uso1_p115_C.
DR InterPro; IPR006953; Uso1_p115_head.
DR Pfam; PF02985; HEAT; 1.
DR Pfam; PF02524; KID; 3.
DR Pfam; PF04871; Uso1_p115_C; 1.
DR Pfam; PF04863; Uso1_p115_head; 1.
SQ SEQUENCE 2042 AA; 233328 MW; 959292DED8EE1158 CRC64;

Query Match 3.9%; Score 200; DB 2; Length 2042;
Best Local Similarity 20.4%; Pred. No. 3.6;
Matches 189; Conservative 150; Mismatches 302; Indels 286; Gaps 43;

QY 20 HNOVFSLVKEPI-----LKQTQASSTISGADYAESSGSKL-----KINFTSGPVD 66
Db 1161 YNDAIEKLNKNDISIASLKEH-SKKVSELD---SGHSLSQDLEAANSRCLTEKQIK 1215
QY 67 DTV--TDLPSDKRTTPKIKONLAKGREQLKAVTETSEKQITSGSQLESLSL 124
Db 1216 EHLSESSNLADQISALEKVG-----ELEASINNAEQE---SNKSRFEKKAEL 1263
QY 125 NKTVPSTSNWEICDFITKGNLTVLGSKSGVEK---LSQTDHLVLPQAADGTQLIOVASF 181
Db 1264 NQNL-----TNLEAKQKAEKRLDLVQEEKAAIEKELAKLQLDDNSK 1307
QY 182 APTDPKKTATAYTSRAGENGESIQDVGDKIINEGEVFNFSYLLKKVITPTGYKHIGQD 241
Db 1308 LET-----EVSELKSDITKPKDEHTIIN---EKLISIKTELSEKKD 1345
QY 242 AP-----VDNKNIAEVNLPESLETI-----SDYAFALAKQI-----D 275
Db 1346 QIENQESKLKDLAKSLDNKEKILVLDLKEKESLETRIKELENDIAYSNSKSEKQTKNEN 1405
QY 276 LPDNLKAI-GELAFPDNQITGKLSLPRQLMRLAERAFKSNHIKTIETFRGNSLKVIGASFP 334
Db 1406 LETKLSKTEKDLSTLNSKFTNETKILKDLI-----SDHEVSI-----SSLKVDLDKVK 1453
```







DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Putative transglycosylase.  
 GN Name=yomi;  
 OS Bacteriophage SPB2.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
 OX NCBI\_TaxID=66797;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Maue C.,  
 RA Karamata D.,  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF020713; AAC13005.1; -.  
 DR F1R, T12796; T12796.  
 DR HSP, P00171; ILSP.  
 DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR002886; Peptidase\_M23B.  
 DR InterPro; IPR008258; SLT.  
 DR InterPro; IPR010090; Tape\_mae\_Tp901.  
 DR InterPro; IPR00189; Transglyc\_AS.  
 DR Pfam; PF01551; Peptidase\_M23; 1.  
 DR Pfam; PF01464; SLT; 1.  
 DR TIGSFAM; TIGR01760; tape\_mae\_Tp901; 1.  
 DR PROSITE; PS00222; TRANSGLYCOSYLASE; 1.  
 SQ SEQUENCE 2285 AA; 252310 MW; FF602C227754B357 CRC64;

Query Match 3.9%; Score 199.5; DB 2; Length 2285;  
 Best Local Similarity 20.9%; Pred. No. 4.4;  
 Matches 215; Conservative 155; Mismatches 356; Indels 303; Gaps 54;

QY 81 EKKIKNLAKPREQELKAVNTES-EXQITSGSQLEQSKSLSLN----- 125  
 DB 771 KKAKDDFQS-QQTNVEAITTKDSTDKLIQYKELQVKYSRSLTDEEQEYLQVTOQL 829  
 QY 126 -KTVPTSNWEICDFTKNTLVGLSKSGVVKLSOT-DHLVLPQAA-----DQTQ 174  
 DB 830 AQTFPAL-----VKGDSQNALTKNKELEKAIENKYLAKQETRDSAKKTFEDASK 885  
 QY 175 LIQVAFAPTPDKTAIAEYTSRAGENGHISQLDVKGKIEINEGVFNSYLLKVTIPTG 234  
 DB 886 EIKKSKDELQYKQ--IADYNDKGRPKWDLIADDDYKVAADKAK-QSMLKAQSDIESG 941  
 QY 235 YKHIGQDAPVDNKNTAEVNLPSLET-ISDYAPFAHALKQIDLPNLKAIGELAFPDNQI 293  
 DB 942 NAKVKDSVLSIANAYSSIDISNTLKTSTSD-----VVKMLKDDLP-BELEKFSSSL 994  
 QY 294 TQKLSLPQLMLAEAPKSNHIKTEPRGNSLKVIGEASFQDNDLSQMLPDGLEKTES 353  
 DB 995 -QKLEKMQ-----KALDSGDEKAFD---NAKK-----DLQSL----- 1024  
 QY 354 EAFTGNPGDDHNNRVLVLTWSKGNPSGLATENTYVNDPKSLWQSSPIDYTKWLEEDFT 413  
 DB 1025 -----ETYSKSDS-----IDVFKSFD--K 1043  
 QY 414 YQKNSVTFPSNGKQKVRKNKLEIPKOHNGVTITEIGNAPRVNDPQNTLRKYDLSEV 473  
 DB 1044 AQKNIKDG--DKSLSSVKSEVG-----DLGETLAEAGNEA---EDFGK-----LKEA 1086  
 QY 474 KLPSIRKIGAPAFQSNLKSPEASDDLEIEKEGAPMNNRIETLEKDKLVITIGDAAPHI 533  
 DB 1087 LDANSVDDIKAAIKEMSDMQP---DSQDVLNGDIPNN-----TKQVAPLND----- 1132  
 QY 534 NHYIAVLPEVQIEGRSAFRQNGANNLIFPMGSKVTKTGEMAFLSNRLHLDLSQKQLT 593  
 DB 1133 -----LLEKMAE-GKS-ISANEANTLI---QKQKELAQAIISIENGVVKINRDEVIQR 1180  
 QY 594 ELPVQAFSD-----NALK-----EVLIPASLKTIRE----- 620  
 DB 1181 KYKLDAYNDMVTYSNKLAKTEVNNNAIKTLNADTLRIDSLKLRKELKMDSEALSLEV 1240  
 QY 621 -----AFKCNHLKQLE-----VASAL-SHIAFNALDDNDGDEQPD 654

DB 1241 KSINNVDADAKKELKLEKMLQPGYNSQIEAMQSVKSALESYI--SASEEATSTQEMN 1298  
 QY 655 NKVVVKTTHNSYALADGEHFIVDPDKLSSTIVD-----LEKILKLEGL-----DYSTLR 704  
 DB 1299 KQALVEAGTSLNWTDDQEKANEBTKTSMYVVDYKKEALEKNAEIDKYNKQVNDYPKYS 1358  
 QY 705 QTTQTQFRD-MTTAGKALLSKSNLRQSEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKK 763  
 DB 1359 Q-----KYRDAIKKEIKALQKKKLMQ--EQAKLLAD-QIKSGNI-----TQYGIIVTST 1404  
 QY 764 ATKNG--QLLERSINKAVLAVNNSAIKKANVKRLEKELDLTLGLVEGK-GPLAQTATVQ 820  
 DB 1405 TSSGTFSSSTGSGYSGKYSYNSAASKYNV-----DPALIAAVIQESGNAKARSVG 1459  
 QY 821 VYLKTEPLPEYVYIGLVYFVDSKGLIYALDMSDTTIGEGQK-DAYGNPILNVDDNEGY 879  
 DB 1460 AMGLMQLMPATAKSLGVNNAYDP-----YQNVMGTKYLAQOLEKFGG--NVEKALAA 1511  
 QY 880 HALAVATLADYEGLD-----IKTIL-----NSKLSQLTSTIRQVPTTAAHYRAGIFAION 928  
 DB 1512 NA-GPGNVIKYGGIPPFKQETQNYVKIMANYSKLSATS-----SIASY----- 1556  
 QY 929 AAAEAEQLLPKPGTHSEKSSSESANGSKDRGLQSNPKTNRGRHSAILPRTGSKG--SPVY 986  
 DB 1557 -----TNSAFRVSSKYQGESLGRSP--HKGTDPAAKAGTAIKSLQSGKV 1601  
 QY 987 GILGYTSTA 995  
 DB 1602 QIAGYSKTA 1610

RESULT 84  
 O31976  
 ID O31976 PRELIMINARY; PRT; 2285 AA.  
 AC O31976;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Yoni protein.  
 GN Name=yomi; OrderedLocusNames=BSU21350;  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377; DOI=10.1038/36786;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M.,  
 RA Klein L.-M., Joris B., Karamata D., Kaehara Y., Klaerr-Blanchard M.,  
 RA Klesin C., Kobayashi Y., Koetter P., Koningstein G., Krogh S.,  
 RA Kunano M., Kurita K., Lapidus A., Lardinois S., Lauber J.,  
 RA Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C.,  
 RA Medigue C., Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S.,  
 RA Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,  
 RA Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,  
 RA Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G.,  
 RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,  
 RA Rose M., Sadale Y., Sato T., Scanlan E., Schleich S., Schroeter R.,  
 RA Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Seror P.,  
 RA Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H.,  
 RA Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P.,  
 RA Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,

Qy	705	QTQTOTFRD-MTTAGKALLSKSNLRQEKEKFLOEAQFFLGRVLDLKAIKAERALKVTKK	763
Dd	1359	Q-----KYRDAIKBEIKALQCKKLMO-EQAQLKDK-QIKSGNI-----TQYGIVTST	1404
Qy	764	ATKMG--QLLERSINKAVLAYNNSAIIKANVKRLKELDLTLGLVECK-GPLAQATWVOG	820
Dd	1405	TSGGTSPSSTGGSYGSYTSINSAASKYNV----DPALTAAVIQQESGFNAKARSGVG	1459
Qy	821	VYLKTPLPLPEYYIGLVNVFDSKGKLIAYALDMSDTTIGEQQ-KDAYGNPILINVDEDNFY	879
Dd	1460	AMGLMQLMPATAKSGLGVNNAYDP-----YQVMWGTKYLAQLEKFCG--NVEKALAAY	1511
Qy	880	HALAVATLAYEGLD-----IKTL---NSKLSOLTSIROVPVTHAYHRAGIFQIAQN	928
Dd	1512	NA-GPGVNVIKYGIPPPFKETONYVYKTMANYSKSLSSATS-----SIASY-	1566
Qy	929	AAAEBEQLLPKPGPHSEKSSSESANSKDRGLQSOPKNTRGHSAIILPRTGSKG--SFVY	986
Dd	1557	-----TNNSAFRVSSKYGQGESGLRSSP--HKGTFPAAKAGTAIKSLQSGKV	1601
Qy	987	GILGYTSVA	995
Dd	1602	QIACYSKTA	1610

Qy		367 GILGIIISVA 993
Dd		1602 QIAGYSKTA 1610
		:
RESULT 85		
QYWE8		
ID	QYWE8	PRELIMINARY; PRT; 3256 AA.
AC	QYWE8;	
DT	01-OCT-2003 (TREMBLrel. 25, Created)	
DT	01-OCT-2003 (TREMBLrel. 25, Last sequence update)	
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)	
DE	Normocyte binding protein 2b.	
OS	Plasmodium falciparum.	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
NCBI_TaxID=5833;		
[1]		
RN	SEQUENCE FROM N.A.	
RP		
RC	SFRAIN=768;	
RA	Rayner J.C., Barnwell J.W., Galinski M.R., Tran T., Huber C.S.,	
FA	Corredor V.;	
RL	Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AY138501; AAN39448.1; -	
SQ	SEQUENCE 3256 AA; 383108 MW; 479FF52584BD8194 CRC64;	
	Query Match 3.9%; Score 199.5; DB 2; Length 3256;	
	Best Local Similarity 19.2%; Pred.No.7.3;	
	Matches 209; Conservative 177; Mismatches 356; Indels 347; Gaps 52;	
Qy	32 ILKQTQASSISGADYAESSGSKLKINETS-----GPVDDTVTDLFSDKRTTPEKIKDN 86	: : : : :           : : : : :           : : : : :
Dd	2026 LVKDCKELRELSTALY-----DLKIQITSVINRENDISNNI-DIVSNKLNEINAIOYN 2077	: : : : :           : : : : :           : : : : :
Qy	87 LAKGPREQELKAVTENPTESEKQI--TSGSOLEQSKE---SLSLNKITVPSTSNWEICDFIT 141	: : : : :           : : : : :           : : : : :
Dd	2078 FEK-----YKEIFDNVVEYKTLDDTKNAIVVKKAELKNVDINKTKE-----DDLDI 2123	: : : : :           : : : : :           : : : : :
Qy	142 KGNTLVGLSGSGVEKUSQTDHLVLPSQAOGTOLIQVASPAFTPDKKTAIARYTSRAGEN 201	: : : : :           : : : : :           : : : : :
Dd	2124 YFNDLDLEKS-----LTLSNMEMEIKTIQV-----NSYNSFSFDIN 2159	: : : : :           : : : : :           : : : : :
Qy	202 GEISQLDVDCG-----ELINGEVFN---SYLLKKVTIPTG----- 234	: : : : :           : : : : :           : : : : :
Dd	2160 KNINDIDKEMKTLPMLDELNLNGHNIDISLYNFIRNIQIKTGNDIKNIREQENDTNIC 2219	: : : : :           : : : : :           : : : : :
Qy	235 -----YKHIGODAFVDNK-----NIAEVLNPESLETISDYAFHALKQI 274	: : : : :           : : : : :           : : : : :
Dd	2220 FEVIQNYYNFIKSDISIFNKYDDHDKVDNYISNNIDVVNKHNSL-LSEH-----VINAT 2272	: : : : :           : : : : :           : : : : :
Qy	275 DLPDN-LKATGEIAFFNQITGKLSUPR-----OLMRLAERAPK-----SNHIKTIEPR--- 322	: : : : :           : : : : :           : : : : :
Dd	2273 NIENIMTSIVEI-----NEDTEWMSLEETQDKLLEYENFKCKGNIIINNMYKIIVHFNKLK 2328	: : : : :           : : : : :           : : : : :



Db 665 RLTLGFAENSLSEIEPEKALEPLTSLNFDLSQNNLALPKTIEKLRALSTIVASRNHIT 724  
 QY 583 HLDLSEQKOLTRI PQAFSDNALKEVLLPASLKTIREFAKKNH-LKQLE-----VASALS 637  
 Db 725 RLNDISFRMLPKSLVDLSTNEISN--LPNGI-----FKONQILKDLFFNNLLTQVE 775  
 QY 638 HIAFNALDDNDGEOFD-----NKVVVVKTH-----HNSYA 667  
 Db 776 EGVFPDVTNLNDFVFNQIKSVSPKVRALIGQHKLTPQKHIAKLEASLDGEKIKYHQAFS 835  
 QY 668 LADGEHFIVDPKLSSTI-----VDLEKILKLE-----GLDYSTLRQTTQTFRDMTTAGKAL 721  
 Db 836 LLDLYW---EQTNSAIDKELSVVEEYQQLQEKSGSTVSLNDMQVDSIVIOLQKKA 892  
 QY 722 LSKSNLRQCEKQFLQEAQFFLGRVDL-DKATAKAEKALVTKKATKNGOLLERSINKAVL 780  
 Db 893 SNGQVTVDEKLLSNDPKDLDLGBFSLDKDPGYRIRKALITIKKFAIQKEHIYLTNSDILV 952  
 QY 781 AYN-----NSAIKKANVRLEK-----ELDLTLGLVEGKGPLAQATWV--Q 819  
 Db 953 AKGPHSHOKDLVENGRLNQQLRDIYIYNASMLKTDLASESNKAINHRVTLVVKK 1012  
 QY 820 GYVLKLTPLPLBEY-----YIGLVNYP-----DKSKLI-----YALDMSDT 856  
 Db 1013 GVSYLEV-----EPRGIKVGKMLGYLGELSYFVDGYQYRDLGKVPGRVTKAEVWSYFTDV 1067  
 QY 857 IGEQKQDAYGN-----PILNVDEDNEG-----YHALAVAT 886  
 Db 1068 TGLPLADRYGKNYPKVLRMKLEQAKQKGLVPLQVFPIMDAISKSGSLQTVFMRLDAS 1127  
 QY 887 LADYEGLDIKTILN-SKLSQLSITSIRV-----PTAAYHRAGIFQAIQNA 930  
 Db 1128 LTTERAKVVKETNNQENSHLSTDLQKPNRQOEKTPSPSPSAATGIANLTDILAKKA 1187  
 QY 931 A-----EASQLLPKPGTHSEKSSSESANSKRGLOSNPKTNKRGHSAIPL 976  
 Db 1188 TQCSQETSKTDDTDKAEKQLQVRDH---QTSIEGKTAKDTTKKSKKHSRNOQSNGE 1244  
 QY 977 RTGSKGSFVYGILGYTSVAL 996  
 Db 1245 ESSRYHLIAGLSSFMIVAL 1264

RESULT 87  
 Q7Q8A9

ID Q7Q8A9 PRELIMINARY; PRT; 1842 AA.  
 AC Q7Q8A9;  
 DT 01-MAR-2004 (TremBLrel. 26, Created)  
 DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)  
 DE AGCP15288 (fragment).  
 GN Name=agCG50148; ORFNames=ENSANGG00000008609;  
 OS Anopheles gambiae str. PEST.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.  
 OX NCBI\_TaxID=180454;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AAA01008944; EAA10144.1; --  
 FT NON TER 1842 1842  
 SQ SEQUENCE 1842 AA; 208361 MW; 978A013DB2ADD5FC CRC64;

Query Match 3.9%; Score 199; DB 2; Length 1842;  
 Best Local Similarity 19.3%; Pred. No. 3.5;  
 Matches 226; Conservative 180; Mismatches 366; Indels 402; Gaps 55;

QY 2 KKHLLKVALTLTTSVVVTH-----NORVFSVLKE-----PILKQTOA 38  
 Db 304 ERSITTEITTEATIVDHTDLRELQENESLRRRLDTVKABAEARDKSDILLRRLAS 363  
 QY 39 SSSISGADYASSGSKSLKINETSFPVDDTVTDLPSDKRTTPEKIK---DNLAGKPR-- 92  
 Db 364 IDTVSNKTAASEALQKQVNEQKOLID---DLQDEKKFLTKLKELSDMSARGGRGI 419  
 QY 93 -----EOLKAVTENTESE-KOITSGSOLEQSKESLSLNTKV-P 129  
 Db 420 EEQLRQKLEQAETCEELMDENEIEKRELNMWETIEEWHDNFREEQADEYASLKKELDQ 479  
 QY 130 STSNWEICDFITKGNLTGLSKSGVEKLSQTDHLVLPQAADGTQLIQVASFAFTPKRT 189  
 Db 480 TTKNCRILSF-----KLKSDR-----RIEQLS-----EKA 506  
 QY 190 AIAEYTSRAGENG-----ISOLDVDGKEIINE-GEVFNISYLLKKVTITP-----GYKH 237  
 Db 507 AL-----GASGDIAAKIKOLE-DELKVSNEVARRLOAELESSTGAPTPTKRTPSLGN 558  
 QY 238 IGQDAFVDNK-----NIAEV--NLPESETISDYAPFAHLAKQIDLPDLNKA 282  
 Db 559 IGKSTSDSKMSRASLRTGGSQEDPAQLRDLQDSLE-----READLREQUKY 606  
 QY 283 IGEAFFDNQITGKL-----SLPRQLMRLAEAFKSNHI-----KTIEFRGNSLK 327  
 Db 607 ABEAEANLRRKSRVEDENDSLVMQLKKWATKATKSSIFSKARKLSPATSRSNADPAP 666  
 QY 328 VIGBASFOO-----NDLSQLMLPDGLEKIESEAFTPGNPDHNNRVV----- 370  
 Db 667 VEKDEGISDEDPAEALRLQLEINQEAMAVLRKMELEHE-----NKHAREQIKDLQE 719  
 QY 371 -LWTKS---CKNPSGLATENTYNP--DKSLWQESPID--YTKWLEEDFTYQKNSVTG 421  
 Db 720 NLATKTKELGRVPSALGSKGVDPLEEKITVMEDEISELRKKLI EKDFERLQ--- 776  
 QY 422 FSNKGLQVKKNKNLEIPKQHNGVTITIGD--NAFRNVDFQNKTLR-----KY 468  
 Db 777 -AEMALAKSGKSLSKTSKSLDALTEQOQDLKRLQVIEQEAIVRAKTQSLEODNEKY 835  
 QY 469 DLEEVKLPSTTRKTGAPAFQSNNLKSPASDDLEI-KEGAPMNNRIE-----TLEKLD 521  
 Db 836 QAEIKKLQTKAKPAGA---TSAADTKKLTDSIEQLEKEKEQLEGLKRIVQESTSOLPM 891  
 QY 522 KLV-----TIGDAAFHINHIYAILVPESVQEIERSAFRQNGANNLIFMGSK 567  
 Db 892 RLAKSPNDMHTYQLRKNVDESELEIDELRALV-----GKA-----GAMNIATLEKE 938  
 QY 568 VKTLGEMAFLSNRLEHLDLSQKQLTEIPVQAFSDNALKEVLLPASLKTIREEAPKQHL 627  
 Db 939 KK-----RLS-AELAEKQOE-----KLITTELSTLKKETLEQOTS 973  
 QY 628 KOLEVASALSIIAFNALDDNDGDFQDNKV-----VVKTHNSVAL----- 668  
 Db 974 KLEBAQRTVERLEDENRRKQDKITLEDKTRVNTMTKTASSKSLLEIQLKAEKEHTG 1033  
 QY 669 --AGEHFIVDPDKLSSTIVLEKILKL-----TEGLDY 700  
 Db 1034 TERDLEKVRKEKTKLDNRI SDLEKELQLSKNAELIKESLEREMAALKSKSATGAEDSES 1093  
 QY 701 STLRTQTTQTPRDMTT-----AGKALLSKSNLRQGEKQKQFLOEAPQFLGR---VDLDK 750  
 Db 1094 SKIQLDLKKQNEELTTEIHQOSRKFEEELQGHETM--EEHLVTVQALAAAREKLOELD 1151  
 QY 751 ---AIAKAEKALVTGKATKNGOLLERSI---NKAVALYNNNSAIKKANVYRLEKLDLITG 804  
 Db 1152 LKNKLLQAE-AIETRLVKENTMGRRLVEVQKLTAAETNDSRYASIE-LEKN----- 1203  
 QY 805 LVEGKGPLAQATMVQGVYLLKTPLPPEYIYGLNVYFDKSKGLIYALDMSDTIGSGQKDA 864  
 Db 1204 -----RMKTALE-----DKQH-----DYEKLAKENEMNA 1227  
 QY 865 YGNPILNVDEDNEGYHALAVATLADYEGLDIKTILNSKLSQLSITSIRQVPTAAYHRAGIFQ 924

```
Db 1228 Y--QVNLKKNEDLR-----KLLDYERIN-----KAQRTLSHNAHLEQ 1266
Qy 925 AIONAAABEQLLPKPGTHSKSSSSSANGSKDR 958
Db 1267 ELKSVYTKLEA-----SEMWKVEVAATRLR 1292

RESULT 88
Q7WS91 PRELIMINARY; PRT; 2186 AA.
AC Q7WS91;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE FmbB-like protein.
GN Name=mip-homologue;
OS Staphylococcus aureus;
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1180;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=PC3, Michigan, New Jersey, St Luke, PC1, Slovenia6096, and
RC SMH2;
RA Wootton M., Avison M.B., Bennett P.M., Howe R.A., MacGowan A.P.,
  Walsh T.R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ567420; CAD98830.1; -
DR EMBL; AJ567421; CAD98831.1; -
DR EMBL; AJ567422; CAD98832.1; -
DR EMBL; AJ567423; CAD98833.1; -
DR EMBL; AJ567424; CAD98834.1; -
DR EMBL; AJ567425; CAD98835.1; -
DR EMBL; AJ567426; CAD98836.1; -
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR011591; Botulinum.
DR InterPro; IPR011439; DUF1542.
DR InterPro; IPR011490; FIVAR.
DR InterPro; IPR005877; Gpos_Y5IRK.
DR Pfam; PF07564; DUF1542; 17.
DR Pfam; PF07554; Y5IRK_signal; 1.
DR Pfam; PF04650; Y5IRK_signal; 1.
DR ProDom; PD001963; Botulinum; 1.
DR TIGRfam; TIGR01167; LPXTG_anchor; 1.
DR TIGRfam; TIGR01168; Y5IRK_signal; 1.
SQ SEQUENCE 2186 AA; 238445 MW; 0CC84836F1868A4F CRC64;

Query Match 3.9%; Score 199; DB 2; Length 2186;
Best Local Similarity 19.2%; Pred. No. 4.4;
Matches 217; Conservative 182; Mismatches 387; Indels 342; Gaps 58;

Qy 20 HNOEVSLVKPILKQTOASSISGADYAESGSKLKNETSGPVDVTDLPSD---- 75
Db 1208 YNAKLAENATPDATNDEKNAINTLNQDROQAIIESIKQANTNAEVDQAATVAENNIDAV 1267

Qy 76 -----KRTTPEKIDNIAKPREQELKAVTENTSEKQITSGSQLEQSK-----SLSLN 125
Db 1268 QVDVVKKQAARDKITAFAK--RIEAVKQTPNATDEEQ--AAVQINQLKQAINQIN 1324

Qy 126 KT--VPSTSNWEI-----CDFTKNTLVGLSKSGVEKLSQTDHLVLPSQAADGTQL 175
Db 1325 QTNQDQVDTTNQAVNAIDNVEAEVVKPKADIADIEKAVKKEQQQIDNSL----DSTDNEK- 1380

Qy 176 IQVASFATPPDKTAIAEYTSRAGNGRISQLDVGKEIINEGEVFNLYLLKVT--Ipt 233
Db 1381 -EVASQALAKEKALAA-IDQAQTNQVQNAATNGVSAIK-----IIQPTKVKKPA 1430

Qy 234 GYKHIGQDAFVNDKNIAEVLNPE-----SLETSIDYAFALH-----KQID 275
Db 1431 AREKINQK---NELRAKINQDKBATEERQVALDKINE--FVQAMTDTITNRTNQVD 1485
```

```
Qy 276 LPDNLKAIGELAPDNOITCKLSLPRLMLAER-----APKSNHIKTIIEPRGSLKV 328
Db 1486 -DTTSQALDSIA-----LVAPEHIVRAAADAVKQYEAKEQIEQAEHATDEKQ 1535

Qy 329 IGBASFQDND-----LSQLMLPDGLEKIESEAFTHGDDHYNNRVVLWTKSGKPSGL 382
Db 1536 VALNQLANNEKLALONINQAVTNDVKRVETNGIATLKG--VQPHIVI-----KPAQQ 1587

Qy 383 ATENTYVNDKSLWQESP-----EIDYTKWL-----BEDFTYQKNSVTGFSN--- 424
Db 1588 AIRATAENQVESI-KOTPHATVDELDEANQLISDTLKQAQOEIENTNQDAADVDVRNQT 1646

Qy 425 KGLQKVK-----RNKN-----LEIPKOHNGVYITEGD--NAPRNV 459
Db 1647 KAIEQIKPVRRKRAALDSIEENKNQLDAIRNTLDTTOBERDVAIDTLANKVINTKNDI 1706

Qy 460 FQNTLRKYDLE-----VKL-----PSTIRKIGAFAFOSNNLKSFEASD-DLEE 504
Db 1707 AQNTNAEVRTETDGDNDNIKVLKPVQVPAARQSVGVKAEQNAL--IDQSDLTSEE 1764

Qy 505 KEGA-----FMNRIETLEKDKLVITIGDAAPHINHIYAILVP-----ESVQEI 549
Db 1765 RLAAKHLVEQALNQAIQDINHADKTAQVNDSDAQNIISKIKPATTVKATALQIQNIA 1824

Qy 550 RSAPFRQGAN-----LIFMGSVKV-----TLGEMAF-----SNRLEHLD 585
Db 1825 TNKINLIKANNEATDEEQNTAIAQVEKELIKAKQIATASAVTNADVAYLLHDEKNEIRE 1884

Qy 586 -----LSEOKOLTEIPVQAFSDNALKEVLLPASLKTIREAPKKNHLKOLEVASALSHI 639
Db 1885 PVNRKASAREQLTTL-----FND-----KQAIENIQATVEE--RNSILAQLQ---NIYDT 1932

Qy 640 AFNALDDNDGDEQFDN--KVVVKTHNSYALADGEHFIVDPDKLSSTIVBLEKILKLEIG 697
Db 1933 AIGQIDQDRSNAQVDKTASLNLQTIHDL-----DVHPFKKPDAAKTINDDLARVTALVQ- 1986

Qy 698 LDYSTLRQTTQTQFRDMTTAGKALLSKNSLRQSEKQKFLQEAQPFGLGRVLDLDAIAKAEK 757
Db 1987 -----NYRKVSNRNK-----ADALK 2001

Qy 758 ALVTKATKNGQLLERSIN--KAVLAYNNSAIKKNVSKLEKELDLTLGLVEGKGLAQ 814
Db 2002 AITALKLOMBELKTARTNADVDVAVLKRFNVALSDIEAVITEKENSLLR--IDN---IAQ 2056

Qy 815 ATMVQGVYLLKTPLPPEYIYGLNVYFDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDE 874
Db 2057 QT-----YAKFKAIATPEQLAKVKVLIDQ-----YVAD-----GNRM--IDE 2091

Qy 875 DNEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAVYHRAGIFQAIQNAABAE 934
Db 2092 D-----ATLN-----DIKQHTQFIVDEILAIR-LPAEA-----TKVSPKEI 2126

Qy 935 QLLPKPGTHSEKSSSSSANSKORGLOSNPKNPKNRGRHSAILPRTGSKG 982
Db 2127 QPAPKVCTPIKKBETHESRKVEK-----LPNTGSEG 2158

RESULT 89
Q99TB0 PRELIMINARY; PRT; 2186 AA.
AC Q99TB0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE SAI577 protein.
GN OrderedLocusNames=SAI577;
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=N315;
RC MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;
EX
```

RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguni A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,  
RT "Whole genome sequencing of methicillin-resistant *Staphylococcus*  
RT aureus".  
RL Lancet 357:1225-1240(2001).  
DR EMBL: AP003135; BAB42845.1; -.  
DR F01; H89960; H89960.  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR011591; Botulinum.  
DR InterPro; IPR011439; DUF1542.  
DR InterPro; IPR011490; FIVAR.  
DR InterPro; IPR005877; Gpos\_Y5IRK.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF07564; DUF1542; 17.  
DR Pfam; PF07564; FIVAR; 1.  
DR ProDom; PD001963; Botulinum; 1.  
DR TIGRfam; TIGR01167; LPXG\_anchor; 1.  
DR TIGRfam; TIGR01168; Y5IRK\_signal; 1.  
DR Complete proteome.  
SQ SEQUENCE 2186 AA; 238448 MW; OCC84836F1868A4F CRC64;

Query Match 3.9%; Score 199; DB 2; Length 2186;  
Best Local Similarity 19.2%; Pred. No. 4.4;  
Matches 217; Conservative 182; Mismatches 387; Indels 342; Gaps 58;

QY 20 HNDVPSLVKPEILKQTOASSISGADYAESGSKLKINETSGPVDVDTLDFSD---- 75  
DB 1208 YNAKLAETINATPDATNDEKNAINTLNQDRQAISIKQANTNAEVDQAATVAENIDAV 1267

QY 76 -----KRTPEKIDNKLAKPREQELKAVTENTESEKQITSGSQLEQSK-----SLSLN 125  
DB 1268 QVDVVKQAARDKITAEVAK--RIEAVKQTPNATDEEK--AAVQINQLKQQAINGQN 1324

QY 126 KT---VPSTSNWEI-----CDFYKGNLVLGSKSGVEKLSQTDHLVLPQAADGTQL 175  
DB 1325 QTNDQVDTTQNAVNAIDNVEAEVWIKPAIAIDIEKAVKEQQQIDNSL---DSTDNEK- 1380

QY 176 IQVASFAFPDPAKTAIAEYTSRAGENGELSOLDVGEKIINEGEVFNLSLLAKVT--IPT 233  
DB 1381 -EVAQALKEKEKALAA-IDQAQNSQVNAATNGVSAIK-----IIQPTKVKRPA 1430

QY 234 GYKHIGQDAFVDNKNIAEYNLPE-----SLETISDYAFALHAL-----KQID 275  
DB 1431 AREKINQKA--NELRAKINQKQKAEAEERQVALDKINE--FVQAMTDTINRTNQVD 1485

QY 276 LPDNLKATGELAFNQITGKLSLPQLMRLAER-----AFKSNHKTIEFRGNSLKV 328  
DB 1486 -DTSQALDSIA-----LVAPHHIVRAAADAVKQYEAQKEIEQAESHATDEEQ 1535

QY 329 ICEASFQND-----LSQLMLPDGLEKLESEAFPTGNDGDDHNNRVNLWTKSGNPSGL 382  
DB 1536 VALNQLANNEKALQINQAVTNNDVKRVETNGIATLKG---VQPHIV-----KPEAQ 1587

QY 383 ATENTVVPDKSLWQESP-----EIDYTKWL-----BEDFTYQNSVTGFSN--- 424  
DB 1588 AIKATAENQVESI-KDTPHATVDELDEANQLISDTLKQAQQEIENTNQDAATVDVNRQTI 1646

QY 425 KGLQKVK-----RNKN-----LEIPKQNGVITIEIGD--NAPRNV 459  
DB 1647 KAEQIKPKVRKRAALDSIEENKNQDLAIRNTLDTQDERDAVIDTLNKIVNTIKNDI 1706

QY 460 FQNTLRKYDLEE-----VKL-----PSTIRKIGAFQSNLKSFEASD-DLBEI 504  
DB 1707 AQNKTNAEVDRTETDGNNDIKVILPKVQVQPAARQSVGVKAEQAQNAL--IDQSDLSSTEE 1764

QY 505 KEGA-----FMNRIETLEKDKLVLTIGDAAFHINHIIVLP-----ESVQETG 549

DB 1765 RLAAKHLVEQALNQALDOINHADKTAQVNOQSDIAQNIISKIKPATTVKATALQIQNTA 1824  
QY 550 RSAPFQNGANN-----LIFMGSKVK---TLGMAFL-----SNRLEHLD 585  
DB 1825 TNKINLIKANNEATDEQNIATIAQVEKELIKAKQOIASAVTNADVAYLLHDEKNEIREIS 1884  
QY 586 -----LSEQKQTEIPVQAFSDNALKEVLLPASLKTIREEAFKNHKLQLEVASALSHI 639  
DB 1885 PVINKASAREQUTTL-----FND---KKQATEANTQATVEE--RNSILAQLQ---NIYDT 1932  
QY 640 AFNALDDNDGDEQFDN--KVVVKTTHNSYALADGSEHFIVDPDKLSTSTIVDLKILKILIEG 697  
DB 1933 AIGQIDQDRSNAQVDTASLNLQTTIDL-----DVHPIKCKPDAEKTINDDLARVLTALVQ- 1986  
QY 698 LDYSTLRQTTQTQTFQDMTGTAGKALLSKNSLRQGEKQKFLQEAQFPLGRVLDLKATAKAEK 757  
DB 1987 -----NYRKVSNNRK-----ADALK 2001

QY 758 ALVTKKATKNGQLLERSIN---KAVLAYNNSAKKANVKRLKELDLTLGLVEGKPLAQ 814  
DB 2002 AITALKQWDEBELKTARTNADVDVAVLKRFNVALSDIEAVITEKNSLLR--IDN---IAQ 2056

QY 815 ATMGVGVYLLKTPPLPEYIYGLNYYFDKSGKLIYALDMSDTIGEGKDAYGNPILNVDE 874  
DB 2057 QT-----YAKPKAIATPEQLAKVKVLIDQ-----YVAD-----GNRM--IDE 2091

QY 875 DNEGYPHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAYHPRAGIFQATONAAAEAE 934  
DB 2092 D-----ATLN-----DIKQHTQFIVDEILAIK-LPAEA-----TKVSPKEI 2126

QY 935 QLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG 982  
DB 2127 QPAPKVCPTIKKEETHESRKEVE-----LPTGUSEG 2158

RESULT 90  
Q6GFU5  
ID Q6GFU5 PRELIMINARY; PRT; 2189 AA.  
AC Q6GFU5;  
DT 05-JUL-2004 (TRENBLrel. 27, Created)  
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)  
DE Putative surface anchored protein.  
GN Name=sac; OrderedLocusNames=SAR1841;  
OS *Staphylococcus aureus* (strain MRSA252).  
OC Bacteria; Firmicutes; Bacillales; *Staphylococcus*.  
RX NCBI\_TaxID=282458;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed-15213324; DOI=10.1073/pnas.0402521101;  
RA Holden M.T.G., Feil E.J., Lindsay J.A., Peacock S.J., Day N.P.J.,  
RA Enright M.C., Foster T.J., Moore C.E., Hurst L., Atkin R., Barron A.,  
RA Bason N., Bentley S.D., Chillingworth C., Chillingworth T.,  
RA Churcher C., Clark L., Cotton C., Cronin A., Doggett J., Dowd L.,  
RA Feltwell T., Hance Z., Harris B., Hauser H., Holroyd S., Jagels K.,  
RA James K.D., Lennard N., Line A., Mayes R., Moule S., Mungall K.,  
RA Ormond D., Quail M.A., Rabinowitch E., Rutherford K.M., Sanders M.,  
RA Sharp S., Simmonds M., Stevens K., Whitehead S., Barrall B.G.,  
RA Spratt B.G., Parkhill J.;  
RT "Complete genomes of two clinical *Staphylococcus aureus* strains:  
RT evidence for the rapid evolution of virulence and drug resistance.";  
RL Proc. Natl. Acad. Sci. U.S.A. 101:9786-9791(2004).  
DR EMBL; EX571856; CAG40832.1; -.  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR011439; DUF1542.  
DR InterPro; IPR011490; FIVAR.  
DR InterPro; IPR005877; Gpos\_Y5IRK.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF07564; DUF1542; 17.  
DR Pfam; PF07554; FIVAR; 1.  
DR Pfam; PF04650; Y5IRK\_signal; 1.



DR TIGRfams; TIGR01167; LPXTG\_anchor; 1.  
DR TIGRfams; TIGR01168; YSIRK\_signal; 1.  
KW Complete proteome.  
SQ SEQUENCE 2189 AA; 238702 MW; 6ECDPFB8D2E28A851 CRC64;

Query Match 3.9%; Score 199; DB 2; Length 2189;  
Best Local Similarity 19.7%; Pred. No. 4.4;  
Matches 201; Conservative 176; Mismatches 359; Indels 284; Gaps 53;

QY 16 SVVTHNQEVSLVKEPIIKQTQASSI-SGADYAESGSKLKINETSGPVDVDTVDLFS 74  
DB 1232 AINTLNQDRQAIES--IKQANTAEVDAQATVAENN-----IDAVQVDVVK 1276  
QY 75 DKRTTPEIKONLAKPREQEELKAVNTESKQITSGSLEOSKE-----SLSLTKT--- 127  
DB 1277 -KQAARDKITAEVAK--RIEAVQTPNATDEEK--AAVNGINQLKDAQFNQINQNTDQ 1332  
QY 128 VPGSTNWEI-----CDFITKGNLTVLGSKSGVEKLSQTHLVLPQAAQDTQLIOVAS 180  
DB 1333 VDATTNQAINADNVEAEVVIKPKAIDIEKAVKEQQQIDNSL---DSTNEK--EVAL 1387  
QY 181 FAFTPDKKTATIABYTSRAGENGESIQLDVQDGKIIINEGEVFNYSLLKVT--IPTGYKH 238  
DB 1388 QALAKEKSKALAA-IDQAQTSQVNAQTNGVSAIK-----IIQPETKIKPAAREKI 1438  
QY 239 GODAFVDNKNIAEVLNPE-----SLETISDYAFAPHALAKQIDLDPNLKAIGELAPF 289  
DB 1439 NOKA---NELRAQINQDKEATAEERQAALDKIND-----LVAKAMTINIDRTNQOVNDS 1490  
QY 290 DNQITGKLSL--PROMLRAER-----AFKSNHIKTIIEFGNSLKVIGEASFQND-- 338  
DB 1491 TQALDDLDIALVTPDHIVRAAADAVKQYEAQKHEIQAEHATDEKQVALNQLANNEKR 1550  
QY 339 -----LSQMLPGLBKISEAFTNGPDHY-----NNRVVLWTSGKRP 379  
DB 1551 ALQNTQAINQANDVNRVSGSIATLKGVEPHIVKPEAQEAIKASADNQV---ESIKDT 1606  
QY 380 SGLATE-----NTYNPDKSLWOESPEIDYTKWLEEDTYQKNSYTGSGNGLQVK--- 431  
DB 1607 PHATTDELDEANQQIN--DTLKQGGQDIDNT---TQDAV--NDVRNQTIKAIEQIKPKV 1659  
QY 432 -----RNKN-----LEIPKQNGVTITEGD--NAPRVNDFONKTLRYD 469  
DB 1660 REKRAALNIDESNNQLDAINTLDTQDERNVAIALNKVNAIKNDIAQKNTNAEVD 1719  
QY 470 LEE-----VKL-----PSTIRKIGAFAPQSNLKSFEASD-DLSEIKEGA----- 508  
DB 1720 QTEADGNNNIKVILPKVQVKKPAARQSVSAKAEQNAL--IDQSLSTEERLAAKHLVEQ 1777  
QY 509 FNNWRIETLELKDVLVTIGDAAPHINHIYAVLP-----ESVQZIGRSAPRQGAN 559  
DB 1778 ALNQAIQINHADKTAQVQNSIDAQNIISKIPATTVKATALQIQIQAINTKINLIKAN 1837  
QY 560 N-----LIFMGSKVK---TIGEMAFI-----SNRLEHL-----LSEQ 589  
DB 1838 NEATDEEQNAALVOVEKELIKAKQIAGATVADVAIYLLHDKGNRIEIRIPVINKKAT 1897  
QY 590 KQLTIPVQAFSDNALKEVLLPASLKTIREAPFKNNHLKQLEVASALSHIAPNALDDNG 649  
DB 1898 EQLATL-----FND---KKQAEANYQATVEE--RNSILAQLQ---NIYDTAIGQIDQDRS 1945  
QY 650 DEQFONKVVVKTHTNSYALADGE-HFIVDPDKLSSTIVDLEKILKIEGLDYSTLRQTQ 708  
DB 1946 NAQVDKATL-----NLQTIHLDVHPHKKPDAEKTIINDDLARVTHLVQ--NRYKVSDRNK 1999  
QY 709 TQFRDMTITAGK-----ALLSKSNLROGEKQKQFQEQAQFGLGRVD----- 747  
DB 2000 ADALAKAITALKQMDDELKARTADNADVAVLKRFNVALGDIIEAVITEKENSILRIDNTAQ 2059  
QY 748 -----LDKAIKAE-----KALVTKKATKNGQLLERS-----INKAVLAYNNSA 786  
DB 2060 QTYAKFAKATPEQAKVAKALIDQYVADGNRMVDEDATINDIKDQTLIIDEILAKLPA 2119

787 -----IKKAN---VKRLKELDLLTGLVBEKGPGLAQATMVQGVYLL 824  
2120 EVIKASPKVQGPAPKVCPTPKEDKQEVKRVVKELP-NTGSEMDLPLKELALITGAALL 2178

RESULT 91  
Q81EB6 PRELIMINARY; PRT; 2651 AA.  
AC Q81EB6;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Ebl-1 protein (Pfl3 0115 protein).  
GN Name=eb1-1; Synonym=PF13\_0115;  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,  
RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.,  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL844509; CAD52344.1; --  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0004872; P:receptor activity; IEA.  
DR GO; GO:0009405; P:pathogenesis; IEA.  
DR InterPro; IPR008602; Duff binding.  
DR Pfam; PF05424; Duff binding; 1.  
SQ SEQUENCE 2651 AA; 305663 MW; 2702399D3979AA68 CRC64;

Query Match 3.9%; Score 199; DB 2; Length 2651;  
Best Local Similarity 19.3%; Pred. No. 5.8;  
Matches 215; Conservative 179; Mismatches 408; Indels 314; Gaps 51;

QY 17 VVTHNQEVSLVKEPIIKQTQASSI-SGADYAESGSKLKINETSGPVDVDTVDLPSDK 76  
DB 1439 IVTEKTLI-----EPTVNILQSTPL-GESHEKSEIIDKDN-TGSEVHTDANTFOYR 1492  
QY 77 RTTPEKIKONLAKPREQEELKAVNTESKQI-----TSGSLEOSKSL--SLNK--- 126  
DB 1493 SSESVEV---ITNKPDEMETTKPPSVIEKEIGESIDSTKQNDQESNIIPLINKNE 1549  
QY 127 ----TVPSTSN-WEICDPITKGNLTVLGSKSGVEKLSQTHLVLPQAAQDTQLIOVASF 181  
DB 1550 EDGVSIPSTRNVMSGSGFVSRNEQII---EKDDKHITDDTTINPSENG-----LKG 1599  
QY 182 AFTPD---KKTAAIABYTSRAGENGESIQ--DVGK-----EINNEGEVFNYSLLKVTI 231  
DB 1600 GEMENDSIKSVTIESPLR-----DVEQMIEPIDGKGNKNNIIGEPQESTTEIRKQMDG 1654  
QY 232 PTGYKHIGQDAFVDNKNIAEVLNPLESLETISDYAFAPHALAKQIDLPD----- 278  
DB 1655 P-----ISNVNPEELHPVAEGSKLSEA-KERSMDDADKGTITEDITV 1697  
QY 279 -----NLKAIHELAP---FDNQITGKLSPLQMLRAER--AFKSNHIKTIIEPRG 323  
DB 1698 EDPNGIGEHQNLKVEHVSQASELNTYNSLDGRTNVEVK-ERLDENPGSIPNDRIITTEHIEL 1756  
QY 324 NSLVIGESAS-----FQNDLSQL-----MLPDGLK 350  
DB 1757 DKEIHELPELDHNGEQEEMIRNEVSDNRMDQISDNETRLQNDHESDSKDEIDK 1816  
QY 351 IESAPTCNPGDDHNNRVVLWTGSKGNPSGLATENTYVNPDKSLMOWESPIDYTKMLER 410  
DB 1817 REMENLEENPNS-----SDSLENPEG-----KEKGKIEHTHSEE 1852  
QY 411 ----DFTYQKNSYTGSGNGLQKVKNNKONLEIPKQNGVTITEIGDNAPRVNDFONKTL 465  
DB 1853 LDSVRDETYK-----YKGIENQITETESVEQQD-----TWIPGNS-KETEVDSRA 1899  
QY 466 RYDLEEVKLPSTIRKIGAFAPQSNLKSFEASDLEIKEGAFMNNRIETLELKDVLVT 525

Db 1900 DMEEEKVVKIKQIITE-----ESEEELEISKQTT--TSHSEKPNIEBQSVN 1943

QY 526 IGDAAPHIN-----HIY-----AIVLPESVQIEGRSAFRONGANNILFMGSKVKTILG 572

Db 1944 IVDKNEINVOIEKNVQNEQEGDPIILAEQKNKTAILENQKYNPNPQSLSHKERTILL 2003

QY 573 ENAFSLNRLE-----HLDLSEKQLTEIPVQAFSDNALKEVLLPASLTKTIIEEAPKQKH 626

Db 2004 EYDDEGSDMTDCLTSELNKNKDSIQIIEASNTDNKLN-----KDITE---NKDD 2052

QY 627 LKQLEVASALSHTAFNALDDNGEQDNKVVVVKTHNS-----YALADGE 672

Db 2053 FSEIEKSGVEIHENGKOLLNKESAEEDVPVQNKIEHDSENAGVIDQYERTDYENKVV 2112

QY 673 HFIVDPDK-----LSSTIVDEKILKILIEGLDYSTILRQTTQTFRDMTTAGKA 720

Db 2113 NFVGLPGKPEERSDVTGRSGVPSVDAFE--KLHEPLD-----2149

QY 721 LLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKAT-----KNGOLLERSIN 776

Db 2150 ----SNEKNKEELNVSEIETYPVGLDILGIDSESSDSVTIYKKPTDFTFVENVHVLSDIK 2205

QY 777 KAVLAYNNSAIKCANVRLKLEKLLDGLVE--GKPLAQATWQGVYLLKTPLPPEY 834

Db 2206 S----LFENEKTVGLNEQKETEKNKTAGMIDVLNNKNIHDDOSVR----LSTSL----YH 2254

QY 835 IGLNVYFDSKGL---IYALDMSDTIGEGQKAYGNPILNVD-----EDNEGVA-----881

Db 2255 KGLD--DOOSGRSLSTALYDKGLDDOQHYLHDSYNNINNIKNKEVCKENCGRNGVSSF 2312

QY 882 LAVATLADYE-----GLDIKTILNSKLSQTSIRQVPTAAYHRAIGFQAIQNAAEAEOL 936

Db 2313 VDTTLRDVKQKEHEETKIENTKIDNTEVDGVDITSDNRSS--FISVDNVDEKKET 2371

QY 937 L-----PKPGTHSEKSSSESANSKDRGLQSNPK 965

Db 2372 IINEKNKDPKPSRKEKQVKKHDDSENNDNNLSNIK 2407

RESULT 92

Q8NWQ6 PRELIMINARY; PRT; 9904 AA.

ID Q8NWQ6

AC Q8NWQ6; STRAIN=WM2;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Ebb protein.

GN Name=ebb; OrderedLocusNames=MW1324;

OS Staphylococcus aureus (strain WM2).

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

OX NCBI\_TaxID=196620;

RN [1]\_

RP SEQUENCE FROM N.A.

RC STRAIN=WM2;

RX MEDLINE=22040717; PubMed=12044378; DOI=10.1016/S0140-6736(02)08713-5;

RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,

RA Nagai Y., Iwano N., Asano K., Naimi T., Kuroda H., Cui L.,

RA Yamamoto K., Hiranatsu K.;

RT "Genome and virulence determinants of high virulence community-

RT acquired MRSA.";

RL Lancet 359:1819-1827(2002).

DR EMBL; AP004826; BAB95189.1; -

DR GO; GO:0016020; C:membrane; IEA.

DR InterPro; IPR011439; DUF1542.

DR InterPro; IPR010983; EF\_Hand\_like.

DR InterPro; IPR011490; FIVAR.

DR InterPro; IPR002988; GA.

DR InterPro; IPR005877; Gpos\_YSRK.

DR InterPro; IPR006530; YD.

DR Pfam; PF07564; DUF1542; 7.

DR Pfam; PF07554; FIVAR; 53.

DR Pfam; PF01468; GA; 48.

DR Pfam; PF04650; YSRK\_signal; 1.

DR TIGRFAMs; TIGR01643; YD repeat 2x; 1.

DR TIGRFAMs; TIGR01168; YSRK\_signal; 1.

KW Complete proteome.

SQ SEQUENCE 9904 AA; 1068489 MW; CE86908F75EA56B2 CRC64;

Query Match 3.9%; Score 199; DB 2; Length 9904;

Best Local Similarity 19.1%; Pred. No. 36;

Matches 223; Conservative 175; Mismatches 451; Indels 316; Gaps 49;

QY 1 MKKHLTKVALTLTTVVVTHNOEVF-----SLVKEPILKQIQ---ASSISGADYA 48

Db 4502 LKEQVQS-ATLLEGVQTVKNSSQTLTAMKGLRDSIANEATIKAGQNYTDASPTNRNEYD 4560

QY 49 ESSGSKSLKINETSFPV--DDTVTDLFSDKRTTPPEKIKDNLAKGPREQELKAVTENTESE 106

Db 4561 SAVTAAKALIINQTSNPTNEPNIIT-----QATSVQTTKE 4594

QY 107 KOITSGSQLEQSKESLSLNTVPSTSNWEICDPIYK---GNVLGLSKSGVE--KLSQT 160

Db 4595 HALANGAQNLAQAKTTAKNNLNLTSINNAQKDALTRSIDGATTVAGVNOETAKATELNA 4654

QY 161 DHLVLPQAAAGCTQLOVASFA-FTPDKKTAI-----AEYTSRACENGELISQLD--- 208

Db 4655 MH-SLQNGINDETQTKQKYLDAEPNKKSAVDQAVNAAKAILTKASGQNVDKAAVEQAL 4713

QY 209 -----VDGKEIINBEVENSYLLKKVTIPTGYKHI--GQDAFVQNK-----NIAEV 252

Db 4714 QNVNSTKTALNGDAKLNKAAKAAKQTLGLT-----HINNAQRNALDNEITQATNVEGV 4767

QY 253 NLPESETISDYAFAPHLAKQIDLDPNL-----KAIGELAFPNQITG 295

Db 4768 NTVKAAQQLDCAMGOLETSIRDKDTLQSQNYQDADAKRTAYSAQVNAATAILNKTAG 4827

QY 296 ----KLSLPRQLMRLAEAFKSNHIKTIYFRGNSLKVIGEASFQNDLS-----340

Db 4828 GNTPKADVERAMQAVQAANTALNGIONLE---RAKQAAANTAITNASDLNTKQEKALKQV 4884

QY 341 ----OLMLPDGLEKIESEAFNGPDDHYNNRVLWTKSGKNPGLATENTVVPD-----392

Db 4885 TSAGVSAANGVEHTATE-----INTAMTALKRAIADKADTKTSGNYVNDANKR 4934

QY 393 -----XSLWOESPEIDYVKWLEBDFYVQKNSVTGFSNGKLOKVKRNKNLEIPKQH 442

Db 4935 QAYDEKVTAAESI VNGTPTTLT---PSDVT---NAAQTQVNTAKTQ-LNGNHNLEVAKQV 4987

QY 443 -----NGVT-----ITEIGDNAFRVVDQNKTLRKYDLEEVKL 475

Db 4988 ANTAIDGLTSLNGPOKAKLKEQVGOATTLPNVQTVRDNA-QTLNTAMKGLRDSIANEATI 5046

QY 476 PS-----TIRK--IGAPAFQSNLKSPEASDDLBEIKEGAF---MN 511

Db 5047 KAGQNYTDASPNRSEYDSAVTAAKAILQQTTPSNNAQETINQAKDQVTKAQALNGQEN 5106

QY 512 NRIETLELKLKLVITGD-----AAPHNHYAIVLPESVQIEGRSAFRQNG 557

Db 5107 LRTAQTNKAKOHLNGLSLDTNAKQAAKQIEGATHVNEV-----TQAQNN 5151

QY 558 ANNLI FMSKVT-IGENAFILSNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKT 616

Db 5152 ADALNTAMTNLKNIGIQDQNTIKQGVNFTDADEAKENAYTNVQAEQILNKAGQENTAKD 5211

QY 617 IREBAFK-----KNHLKQLEVASALSHTAFNALD-----DNDGDGEQFNKVVVVKTHNSY 666

Db 5212 NVESALQNVQRAKNELNGQNVANNAKTAKAKNNLNTSINNAQKALKSQI-----5262

QY 667 ALADGEHFIVDPKLSSTIVLEKIL-KLIEGLD-----YSTLRQTTQTFRDMT 715

Db 5263 ---EGATTVAGVQVSTTASELNTAMSNLQGINDEAATKAAQKYTDADRDKQYANDAV 5319

QY 716 TAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKATKNGQLLERSI 775

Db 5320 TAAKTLLDKT-AGTNNENKAAVEQA---LQRVNTAKTALNGDARLNEAKNTAKQV-----5370







```
RESULT 96
Q862A8 PRELIMINARY; PRT; 1169 AA.
AC Q862A8;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Kinesin.
GN Name=KLP3;
OS Cochliobolus heterostrophus (Drechslera maydis).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochliobolus.
OX NCBI_TaxID=5016;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22627967; PubMed=12742059; DOI=10.1016/S1087-1845(03)00022-7;
RA Schoch C.L., Aist J.R., Yoder O.C., Gillian Turgeon B.;
RT "A complete inventory of fungal kinesins in representative filamentous
ascomycetes.";
RL Fungal Genet. Biol. 39:1-15(2003).
DR EMBL; AY230428; AAO59290.1; -.
DR HSSP; P17119; 1F9T.
DR GO; GO:0005875; C:microtubule associated complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
DR ATP-binding; Microtubule; Motor protein.
KW ATP-binding; Microtubule; Motor protein.
SQ SEQUENCE 1169 AA; 129138 MW; 783AF904F844DDCE CRC64;

Query Match 3.9%; Score 197.5; DB 2; Length 1169;
Best Local Similarity 19.9%; Pred. No. 2.1;
Matches 198; Conservative 155; Mismatches 344; Indels 297; Gaps 48;

QY 45 ADYAESSGK-SKLKINETSFGPDDVTVDLPSDKRTTPEKIKDN-----LAGPBEQ 94
DB 164 ADBIENSVCSPFEL-----YNEELRLLSPDDATKLKYDNDSSKSGSTTLTVQGMSEC 217
QY 95 ELKAVTENTESKQITSGSQLESLSLNKTPSTNWEICDFITKGNLTVLGSKGV 154
DB 218 HLKTASEGI---KLLANGSHKQVATKCNLDSSRSHVTFTVYIKR-----TTEDQG 268
QY 155 EKLSTQDHLVPSQADGTQLIQVAFAPTPDKKTAIAEYTSRAG-ENGEISQLDVGKE 213
DB 269 EYLS-----AGKLNLDLAG-----SENIQRSGAENKRAAEAGLINKS 306
QY 214 IINEGVNSVLLKKVTIP---TGKHTIGODA-----FVDKNIAEVLNLPESLETIS 262
DB 307 LTLGRVNALVERSHPIPYRESKLTLLQDSLOGRTKTCIATLSPAKSNLEETISTL- 365
QY 263 DYAFALHAK---QIDLPNLKAIGELAFFDNOITGKLSLPROLMRLAERAFKSNHIXTI 319
DB 366 DYAFPAKIRKPNQVQNPINKTL--LKEYTWEIE---KLKSELATQR-----NGVILT 416
QY 320 EFRGNLSKVIQSGFQNDLSQMLPDGLEKIESEAPFGNPGDDHNNRVVLMTKSGRNP 379
DB 417 QENYEITITSES-----RRILSEQRERLET-----NEVNLRTKV-EDL 455
QY 380 SGLATENVNPKSLMWESPIDYTKWLEEDFTYQKNSVTGFSNKGKQKVRN---KNLE 437
DB 456 FKLTATKFTQLKKDNE--QTQALDGTGKILE-----KTEIV-----LNHTRQNLDDETE 502
QY 438 IPKQNGV--TITEIGDAPRNVDPQNTLRYKYLEEVKLPSTIRKIGAFAPQSNLKSFP 495
DB 503 LRKAHQKTEELAEVGRDMMSTL-----GKTTSAID-----GLRSKIKR--KSELQSQNRN 553
QY 496 BASDDLEBIKEGAFMNNRIETLELKDVLVTIGDAAFHINHIYAIVLPSVQEIQRSAFRQ 555

Db 554 NSS-QTVVDVTRLVVEDRIE--EFHQEQEQLMDA-----ISERMQTFVRDELEK 599
QY 556 NGANNLIIPWGSKVTKLGEWAFLSNPLEHLDLSEQQLTEIPVQ-AFSDNALKEVLLPASL 614
DB 600 LGAS-----OSFLOEKMEAYQTS-----AEVNGQTAOARDHNMNEVL--EEI 639
QY 615 KTIREEAFKKNHLKOLEVASALSALSHIAFNALDNDGDEQFNDKVVVKTHHNSYALADGEHF 674
DB 640 KTLREEVTKIGAGLDLSAAETISANIITELDA---FHTQV-----HSSYA-SLGRDF 690
QY 675 IVDPKLSTTVDLKILKLEGLDYSTLRTOTQTFQDMTTAGKALLSKSNLRQ----- 730
DB 691 -----KTTFDLWK-----DLNDQQAENERLHQVVVEANTALTEAHKASEGQLAK 735
QY 731 ---EKQFLOEAOFFLGRVLDLKAIAKAEKALVTTKATKNGQLLERSINKAVLAYNNSA 786
DB 736 LVDEKQKASDRQELLAQI-----TALMTANA----- 763
QY 787 IKKANVRLKELEKELDLLTGLVEGKGPLAQATMVQGVYLLKTLPLPLPEYVIGLVNVPDKSGK 846
DB 764 --DAQEKELEERLSAVRGEIAAANTAPEAKQ-----SAYSEGVNAWSDKSRD 808
QY 847 LIYALDMS-DTIGEGQKDAYGNPI-----LNVLD----- 873
DB 809 ILAGVSKSRRTVTKIKSDFAVSVCRFPTSHANVDQAATEHSTLIKETTTSVHSTVKT 868
QY 874 -----EDNEGYHA---LAVATLADYEGLDIKTI---LNSKLS 904
DB 869 VEAQWAHLDTQLQSLDDIVSRIRBONNAHHAHTTSLAALSTVGASYSIGEHLASSFD 928
QY 905 QLTSTIRQVPTAAVHRAGIFQAIQNAABAEQLLP 938
DB 929 RVQSESEMSA--HTAGLKETLPSLSAESDIRAP 960

RESULT 97
Q07380 PRELIMINARY; PRT; 1790 AA.
AC Q07380; P89892;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE US01 protein.
GN Name=US01;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloecker H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z74106; CAA98621.1; -.
DR PIR; S67593; S67593.
DR HSSP; P01106; INKP.
DR SGD; S00002216; US01.
DR GO; GO:0005625; C:soluble fraction; IDA.
DR GO; GO:0006888; P:ER to Golgi transport; IMP.
DR GO; GO:0006481; P:protein complex assembly; IDA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR02017; Spectrin.
DR InterPro; IPR006955; US01_p115 C.
DR InterPro; IPR006953; US01_p115 C.
DR Pfam; PF04871; US01_p115 C; 1.
DR Pfam; PF04869; US01_p115_head; 1.
SQ SEQUENCE 1790 AA; 206450 MW; 90062544F5A52EE CRC64;

Query Match 3.9%; Score 197.5; DB 2; Length 1790;
```





```
Db 1842 NRMEQISRDNETROLNQBHSDKDEIIDKREMLEENLEPNSS-----SDSLENP 1892
Qy 380 SGLATENTYVNPDKSLWQSPEDYTKWLEE-----DFTYQKNSVTGFSNKGLOKVRNK 434
Db 1893 EG-----KKGKIEHTSHSEELDSVRDEYK-----YKGIENQITET 1929
Qy 435 NLEIPKQHGVTITEIGDGNFRNVDPQNTLRLKYDLEEVKLPSTIRKIGAFQSNLKS 494
Db 1930 EIESVEQD-----TNIPGNS-KETEVDNDRADMEBEKVKIKIITE----- 1971
Qy 495 FASDDLEIEKGAFAWNNRIETLEKDLVTGDAAFHIN-----HIY-----AIVL 541
Db 1972 -ESEELEISKDTT--TSHSEKPSIEEQSVNIQDSKNEINVQIEKNVQNEQEGDPILL 2028
Qy 542 PESVQIEGRSAFRONGANNLIFMGSKVKTIGEMAFLSNRLE-----HLDLSEQQLTEI 595
Db 2029 EQNKVIALENOKNKYNPNPSQSHKERTILLEVDLEGMDTNTCLTSELNKNKCSIQII 2088
Qy 596 PVQAFSDNALKEVLLPASLKTIRREAPKKNHLKQLEVASALSHIAFNALDDNDGDEQFND 655
Db 2089 PEASNTDNKLN-----KDITE---NKDDFSEIEKSVGEIHENGKOLLNKESAESDDV 2137
Qy 656 KVVVKTHERNS-----YALADGEHPIDVDPK-----LSSTIVDLE 689
Db 2138 PVQNKIEHDSNAGVIDOYEYRTDYENKVNPFGLPGKPEBERSDVTTRGSPVSDAPE 2197
Qy 690 KILKIEGLDYSLTRQTOTQFRDWTAGKALLSKSLRQGEKQFLQEAQFPLGRVLD 749
Db 2198 K--KLHEPDL-----SNEKNKELNVSKIEIYFVGLDILG 2230
Qy 750 KAIKAEKALVTKKAT-----KNGQLLERSINKAVLAYNNNSAIKANVKRLEKELDLITGL 805
Db 2231 IDSSESDSVTIYKRPDTFVENVHVLSDIKS---LFENEKTVGLNEQKKEKNIAGCM 2287
Qy 806 VE--GKGPLAQATMGVGLLKTPLPLPYIYGLNVYFDKSKL-----IYALDMSDTIGR 860
Db 2288 IDVLNKNKIDDDQSVR---LSTSL-----YHKGLED--DDQSGRLSTALYDKGLDDQHYL 2337
Qy 861 QKDAYGNPILNVED-----NEGHYHALA-VATLADYE-----GLDIKTILSKLSQ 905
Db 2338 HDNSNNNNNNINKNEVQCKENGCRNENSVFVDTTLTRDVQKQKHEETKLIENTKIDN 2397
Qy 906 LTSIRQVPTAAVHRAGIFAQIAQNAABARQLL-----PKPGTHSEKSSSESANSKOR 958
Db 2398 TEVDGVDITSDNRSS-FLISVDNYDEKKETIINEKKNDPKSRKEEKDVKHDDSENDNN 2456
Qy 959 GLQSNPK 965
Db 2457 NLSSNIK 2463
RESULT 99
Q7P2P7
ID Q7P2P7 PRELIMINARY; PRT; 197 AA.
AC Q7P2P7;
DT 01-WAR-2004 (Tremblrel. 26, Created)
DT 01-WAR-2004 (Tremblrel. 26, Last sequence update)
DE Hypothetical cytosolic protein.
GN Name=FN1342;
OS Fusobacterium nucleatum subsp. vincentii ATCC 49256.
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium
OC NCBI_TaxId=209882;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49256;
RA Karpatal V., Ivanova N., Anderson I., Reznik G., Bhattacharyya A.,
RA Gardner W.L., Mikhailova N., Larsen N., D'Souza M., Walunas T.,
RA Haseikorn R., Overbeek R., Kyrpides N.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
```

```
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABF01000055; EAA24104.1; -.
KW Hypothetical protein.
SQ SEQUENCE 197 AA; 22444 MW; 3ED165BAEE414D32 CRC64;
Query Match 3.9%; Score 197; DB 2; Length 197;
Best Local Similarity 26.6%; Pred. No. 0.19;
Matches 54; Conservative 43; Mismatches 92; Indels 14; Gaps 4;
Qy 402 IDYTKWLESDFTYQKNSVTGFSNKGLOKVK--RNKNLEIPK-OHNGVTITEIGDGNFRNV 458
Db 1 MDQNIWEYDDDFPKGDELKGMTQKQKVKLEKTDLVIPELTPDGLPLKKIADNAF--- 57
Qy 459 DFQNTKLRKYDLEEVKLPSTIRKIGAFQSNLKSFEASDDLEIEKGAFAWNNRIETLE 518
Db 58 -----YRGLTSTVIVTNTVESIGYDAFGVCKLKEVKLPKALVNIEGFAFRNKLTKVE 110
Qy 519 LKDLVTIGDAAFHINHIYAIVLPSVQIEGRSAFRONGANNLIFMGSKVKTIGEMAFLS 578
Db 111 FGSKVKRLPSPSSFAMNELAEVTFPETLEYIGASAFYKKNFETVSPKSVTK-IDMYAPRK 169
Qy 579 NRLEHLDLSEQQLTEIPVQAPS 601
Db 170 NNHKVEVANSVDLHKFAFEPT 192
RESULT 100
Q8GC27
ID Q8GC27 PRELIMINARY; PRT; 897 AA.
AC Q8GC27;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Internalin B, i-inlB2 protein precursor.
GN Name=i-inlB2;
OS Listeria ivanovii.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OC NCBI_TaxId=1638;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19119;
RA Dominguez-Bernal G., Gonzalez-Zorn B., Garrido P.,
RA Vazquez-Boland J.A.,
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ271621; CAC81246.1; -.
DR HSSP; P25146; 106V.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR007092; LRR SDS22.
DR Pfam; PF00560; LRR 1; 15-
DR PRINTS; PR00019; LEURICHRPT.
KW Signal.
FT SIGNAL 1 33 Potential.
FT CHAIN 33 897 i-InlB2 protein.
SQ SEQUENCE 897 AA; 100397 MW; 4CE2608B29DBF314 CRC64;
Query Match 3.9%; Score 197; DB 2; Length 897;
Best Local Similarity 17.8%; Pred. No. 1.6;
Matches 164; Conservative 172; Mismatches 373; Indels 214; Gaps 36;
Qy 98 AVTENTESEKQITSGSQLEQSKESLSLNKTVPSSTSNWBIQDPITKGNLTVG-LSKSGVEK 156
Db 20 AMCVNTNLEIKVQRAAS-----ISNPMPIQIIPDTNLAEVWKSLRKNSTVDVVSQELDQ 75
Qy 157 LSQTDHLVLPQAADGTQLIQVAFPTPDKKTAETVTSRAGENGEISQLDVGKEIN 216
Db 76 VESINAMKNIKSIQGLQHLNLTQIYLSNDQIQDISYLTNL---NKLEEIYLSGNQIKD 132
Qy 217 EGEVPSVLLKKVTIPTQYKIHGDQAFVDNKNIAEVLNLPESLETISDYAFALHAKQIDL 276
Db 133 IGHLANLNKLEKI-----FLOGNQLTDINLPAGLSLNKTLVLNNOIKICN 179
Qy 277 PDNLKAIGELAFFDQNIQTKLSLPQLMRLAERAFKSNHKTIEPRGNSLVKIGEASFQD 336
```

Db	180	LEKSKLNVYLOQNT-DISI-AGUSNLWILDSNNQIKGINOLAN-LNKINELYLEG	236
Qy	337	NDSLSQMLPDGLEKIESEAFGNPGDDHNNRNVVLWTKSGKNPSGLATENTVYNPKSLW	396
Db	237	NQLTDISVLAGLSNLKTL-----DLNNRI-----KDIRTLSTLVLENLLM	278
Qy	397	QESPEIDYTKWLEEDFYQKNSVTGFSNGKLGQKVRKNKNLIPKQHNQVTTTEIGDNAPR	456
Db	279	NNNQLINI-----NHLSSLLKLLSPFNGRVTTISSVAKLTNLTET-DCSEN	325
Qy	457	NVDFONKTLRKYDLEEVKLPSTIRKIGAFAFQSNNLKSFEASDDLEETKEGAFNNRRET	516
Db	326	QVDNINSIAKLNTLTGLT-----EGNKVKDLSPLAQLTNLTGLNFRQNLND	373
Qy	517	LELKDKLVITGDAAAPHINHIYAVLPBSVQBIGRSAPFQNGANNLIFMGSKVKTIGEMAF	576
Db	374	ISIDEKLPNLDLSLAFDKNKVSDVSILAKLPKLTLYLFNDNQVTNIDSL-AKLPHLVGVDF	432
Qy	577	LSNRLEHLDLSEQOLTEIPQVAF---SDNALKEVLLPASLKTREEAPFKNNHLKQLEVA	633
Db	433	SGNKVSI-----KALTNLTKLRFNLNANGNCIQDIALRGUTJUELKLRNRINDI---	484
Qy	634	SALSHIAFNALDDND-GDEQFDNKV-----VVKTHNSYAL--ADGEHFIVDPD	679
Db	485	SPL--IWLNNIDELDSNQAFINRPIDFQVNVTPINIVKDTGTGLIAPNSSDKVTYNPP	542
Qy	680	KLSSTIVDLKILKLIBGLDYSTLRQTQTQFR-----DMTTAGKALLSKSNLRQ	729
Db	543	NLTWNLPTYKK--EVCVTFNQSVMLGQATSFSGQVQPFNKLDVVKSNKAITAYGRVKP	600
Qy	730	GE-----XQ-----KFOEA-----QFFELGRVLDLDAKIAKA	755
Db	601	GKTSVTWVTPQRTKDTQKQVGMVSAYTGKHLIRLEAQTTSGIYYQFSVGGKTIGHWEAQA	660
Qy	756	-----EKALVTKKAT-----KNGQ-----LLERSINKAVLAYNNSAIKANYKRLK	797
Db	661	LGIIFYQSSMEKKATGARVTVRGKEGFHFKLPVIDAADGGTLA-----KWKGRLLII	713
Qy	798	ELDLITG-----LYVEKG-----PL---AQATWVQGVYLLKTPLPPEYII	835
Db	714	QREVTIGKEKWLLOGVGWVKATNLTTVLYNKPMTGYAKVKLAKGVVMSNPYNTIGYKV	773
Qy	836	--GLNVFVFDKSGKLIYALDMSD-----TIG--EG-----OKDAYGNPI	869
Db	774	VASLSRYTGKVLRIILREAQTSNGLYYQYVRVGKKTIGVWEAKNVAFYFKPIMEKKYKTRF	833
Qy	870	LNVDNEDGYHALAVATLADYEG	892
Db	834	VKVEKQKQAYYSMPVADAAISRG	856

Search completed: August 28, 2005, 10:45:28  
Job time : 236 secs



103 7 0.7 264 2 T23866 hypothetical prote  
104 7 0.7 268 2 T15693 hypothetical prote  
105 7 0.7 269 2 D69996 amino acid ABC tra  
106 7 0.7 269 2 AE1368 amino acid ABC tra  
107 7 0.7 269 2 AF1737 amino acid ABC tra  
108 7 0.7 269 2 B69381 hypothetical prote  
109 7 0.7 269 2 T22739 hypothetical prote  
110 7 0.7 270 2 E69996 amino acid ABC tra

## ALIGNMENTS

## RESULT 1

T51594  
Glutathione transferase (SC 2.5.1.18) AtGST 10 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 09-Jul-2004  
C:Accession: T51594  
R:Dixon, D.P.; Cole, D.J.; Edwards, R.  
Plant Physiol. 119, 1568, 1999  
A:Title: Identification and cloning of AtGST 10 (Accession Nos. AJ131580 and AJ132398),  
A:Reference number: Z25404  
A:Accession: T51594  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-245 <DIX>  
A:Cross-references: UNIPROT:Q9ZRT5; EMBL:AJ131580; PDB:CAA10457.1  
A:Experimental source: variety Columbia; seedling; whole plant  
C:Genetics:  
A:Gene: gst10  
C:Keywords: transferase

Query Match 0.9%; Score 9; DB 2; Length 245;  
Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAAAEQLL 937

DB 133 AAAAAEQLL 141

## RESULT 2

J90176  
Orphan G protein-coupled receptor precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 03-Jul-1998 #sequence\_revision 10-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: J90176  
R:McDonald, T.; Wang, R.; Bailey, W.; Xie, G.; Chen, F.; Caskey, C.T.; Liu, Q.  
Biochem. Biophys. Res. Commun. 247, 266-270, 1998  
A:Title: Identification and cloning of an orphan G protein-coupled receptor of the glyco  
A:Reference number: J90176; MUID:98308104; PMID:9642114  
A:Accession: J90176  
A:Molecule type: mRNA  
A:Residues: 1-907 <MCD>  
A:Cross-references: UNIPROT:O75473; GB:AF062006; NID:93366801; PDB:AC28019.1; PDB:9336  
C:Comment: This protein is a receptor for a novel class of glycoprotein ligands.  
C:Genetics:  
A:Gene: HG38  
A:Map position: 12q22-23  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:562-583/Domain: transmembrane #status predicted <TM1>  
F:594-616/Domain: transmembrane #status predicted <TM2>  
F:639-660/Domain: transmembrane #status predicted <TM3>  
F:681-701/Domain: transmembrane #status predicted <TM4>  
F:725-744/Domain: transmembrane #status predicted <TM5>  
F:768-791/Domain: transmembrane #status predicted <TM6>  
F:803-824/Domain: transmembrane #status predicted <TM7>

Query Match 0.9%; Score 9; DB 2; Length 907;  
Best Local Similarity 100.0%; Pred. No. 5.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVOAF 600  
DB 174 LTEIPVOAF 182

## RESULT 3

T46320  
Hypothetical protein DKFZp434H0512.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004  
C:Accession: T46320  
R:Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23035  
A:Accession: T46320  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-116 <AAA>  
A:Cross-references: UNIPROT:Q9NSZ8; EMBL:ALJ137620  
A:Experimental source: adult testis; clone DKFZp434H0512  
C:Genetics:  
A:Note: DKFZp434H0512.1

Query Match 0.8%; Score 8; DB 2; Length 116;  
Best Local Similarity 100.0%; Pred. No. 9.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 930 AAAAAEQLL 937

DB 83 AAAAAEQLL 90

## RESULT 4

T08175  
Chlorophyll a/b-binding protein homolog LI818r-1 precursor - Chlamydomonas reinhardtii  
C:Species: Chlamydomonas reinhardtii  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: T08175  
R:Guertin, M.  
submitted to the EMBL Data Library, January 1996  
A:Reference number: Z16375  
A:Accession: T08175  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-253 <GUE>  
A:Cross-references: UNIPROT:P93664; EMBL:X95326  
A:Experimental source: strain CC620  
C:Genetics:  
A:Gene: LI818r-1  
C:Superfamily: chlorophyll a/b-binding protein

Query Match 0.8%; Score 8; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 18;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 LPDNLKAI 283

DB 246 LPDNLKAI 253

## RESULT 5

E71613  
Hypothetical protein PF0490c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: E71613  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: E71613  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

Query Match 0.9%; Score 9; DB 2; Length 907;  
Best Local Similarity 100.0%; Pred. No. 5.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

A:Molecule type: DNA  
 A:Residues: 1-284 <GAR>  
 A:Cross-references: UNIPROT:O96191; GB:AE001398; GB:AE001362; NID:g3845197; PIDN:AA7188  
 A:Experimental source: clone 3D7  
 C:Genetics:  
 A:Gene: PFB0490C

Query Match 0.8%; Score 8; DB 2; Length 284;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 EVNLPESL 258  
 |||||  
 Db 269 EVNLPESL 276

RESULT 6.  
 C95961  
 probable sugar uptake ABC transporter periplasmic solute-binding protein precursor Smb21  
 C:Species: Sinorhizobium meliloti  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 09-Jul-2004  
 C:Accession: C95961  
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
 A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endo  
 A:Reference number: A95842; MUID:21396508; PMID:11481431  
 A:Accession: C95961  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-296 <KUR>  
 A:Cross-references: UNIPROT:Q26606; GB:AL591985; PIDN:CAC49355.1; PID:g15140841; GSPDB:G  
 A:Experimental source: strain 1021, megaplasmid pSymb  
 R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: Smb21377  
 A:Genome: plasmid

Query Match 0.8%; Score 8; DB 2; Length 296;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 751 ATAIAEKA 758  
 |||||  
 Db 98 ATAIAEKA 105

RESULT 7.  
 A87258  
 hypothetical protein CC0074 [imported] - Caulobacter crescentus  
 C:Species: Caulobacter crescentus  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
 C:Accession: A87258  
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
 n, J.; Emoleeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A:Title: Complete Genome Sequence of Caulobacter crescentus.  
 A:Reference number: A87249; MUID:21173698; PMID:11259647  
 A:Accession: A87258  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-304 <STO>  
 A:Cross-references: UNIPROT:Q9ABZ6; GB:AE005673; NID:g13421171; PIDN:AAK22061.1; GSPDB:G  
 C:Genetics:  
 A:Gene: CC0074

Query Match 0.8%; Score 8; DB 2; Length 304;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 881 ALAVATLA 888  
 |||||  
 Db 131 ALAVATLA 138

RESULT 8  
 T37695  
 hypothetical protein SPAC14C4.10c - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C:Accession: T37695  
 R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
 submitted to the EMBL Data Library, August 1997  
 A:Reference number: Z21737  
 A:Accession: T37695  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-329 <DEV>  
 A:Cross-references: UNIPROT:O13717; EMBL:Z98596; PIDN:CAB11203.1; GSPDB:GN000066; SPDB:SF  
 A:Experimental source: strain 972h-; cosmid c14C4  
 C:Genetics:  
 A:Gene: SPDB:SPAC14C4.10c  
 A:Map position: 1  
 A:introns: 140/1

Query Match 0.8%; Score 8; DB 2; Length 329;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 798 ELDLITGL 805  
 |||||  
 Db 4 ELDLITGL 11

RESULT 9  
 G97025  
 probable membrane protein [imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
 C:Accession: G97025  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
 J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: G97025  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-429 <KUR>  
 A:Cross-references: UNIPROT:O97KAL; GB:AE001437; PIDN:AAK78994.1; PID:g15023928; GSPDB:G  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC1018

Query Match 0.8%; Score 8; DB 2; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 841 FDKSGKLI 848  
 |||||  
 Db 239 FDKSGKLI 246

RESULT 10  
 B84048  
 argininosuccinate lyase argh [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004



C:Accession: B84048  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Maui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 C:Accession: B84048  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-458 <STO>  
 A:Cross-references: UNIPROT:Q9K821; GB:AP001518; GB:BA000004; NID:gl0175792; PIDN:BA8069  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: argH  
 C:Superfamily: argininosuccinate lyase

Query Match 0.8%; Score 8; DB 2; Length 458;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 KAIAKAEK 757  
 DB 447 KAIAKAEK 454  
 |||||

RESULT 11  
 C72658  
 Hypothetical protein APE0691 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
 C:Accession: C72658  
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
 DNA Res. 6, 83-101, 1999  
 A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
 A:Reference number: A72450; MUID:99310339; PMID:10382966  
 A:Accession: C72658  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-478 <RAW>  
 A:Cross-references: UNIPROT:Q9YE79; DDBJ:AP000060; NID:95104188; PIDN:BA79667.1; PID:d1  
 A:Experimental source: strain K1  
 C:Genetics:  
 A:Gene: APE0691  
 C:Superfamily: Aeropyrum pernix hypothetical protein APE0691

Query Match 0.8%; Score 8; DB 2; Length 478;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 LKVIGEAS 333  
 DB 406 LKVIGEAS 413  
 |||||

RESULT 12  
 JQ2248  
 Beta-amylase (EC 3.2.1.2) - rye  
 N:Alternate names: 1,4-alpha-D-glucan maltohydrolase  
 C:Species: Secale cereale (rye)  
 C:Date: 03-May-1994 #sequence\_revision 25-Oct-1996 #text\_change 09-Jul-2004  
 C:Accession: JQ2248  
 R:Sadowski, J.; Rorat, T.; Cooke, R.; Delseny, M.  
 Plant Physiol. 102, 315-316, 1993  
 A>Title: Nucleotide sequence of a cDNA clone encoding ubiquitous beta-amylase in rye (Se  
 A:Reference number: JQ2248; MUID:94151427; PMID:8108499  
 A:Accession: JQ2248  
 A:Molecule type: mRNA  
 A:Residues: 1-503 <SAD>  
 A:Cross-references: UNIPROT:Q08335; GB:Z11772; NID:9393449; PIDN:CAA7817.1; PID:G93450  
 C:Function:  
 A:Description: hydrolyzes 1,4-glycosidic linkages of starch, removing maltose units succ  
 C:Superfamily: beta-amylase  
 C:Keywords: glycoprotein; glycosidase; hydrolase; monomer; polysaccharide degradation

F:2-503/Product: beta-amylase #status predicted <MAT>  
 F:184,378/Active site: Glu #status predicted  
 F:249,338/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 0.8%; Score 8; DB 1; Length 503;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811  
 DB 55 GLVEGKGP 62  
 |||||

RESULT 13  
 S00222  
 beta-amylase (EC 3.2.1.2) - barley  
 N:Alternate names: 1,4-alpha-D-glucan maltohydrolase  
 C:Species: Hordeum vulgare (barley)  
 C:Date: 30-Jun-1989 #sequence\_revision 25-Oct-1996 #text\_change 09-Jul-2004  
 C:Accession: S00222; JX0299; PC2019  
 R:Kreiss, M.; Williamson, M.; Buxton, B.; Pywell, J.; Hejgaard, J.; Svendsen, I.  
 Eur. J. Biochem. 169, 517-525, 1987  
 A>Title: Primary structure and differential expression of beta-amylase in normal and mut  
 A:Reference number: S00222; MUID:88082785; PMID:2446870  
 A:Accession: S00222  
 A:Molecule type: mRNA  
 A:Residues: 1-535 <KRR>  
 A:Cross-references: UNIPROT:P16098; EMBL:X52321; NID:gl8917; PIDN:CAA36556.1; PID:gl8918  
 A>Note: parts of this sequence, including the amino end of the mature protein, were dete  
 R:Yoshigi, N.; Okada, Y.; Sahara, H.; Koshino, S.  
 J. Biochem. 115, 47-51, 1994  
 A>Title: PCR cloning and sequencing of the beta-amylase cDNA from barley.  
 A:Reference number: JX0299; MUID:94245663; PMID:8188635  
 A:Accession: JX0299  
 A:Molecule type: mRNA  
 A:Residues: 1-232, 'A', 234-346, 'S', 348-526, 'M', 528-535 <YOS>  
 A:Cross-references: GB:D21349; NID:9415599; PIDN:BAA04815.1; PID:g464145  
 A:Accession: PC2019  
 A:Molecule type: protein  
 A:Residues: 129-141;152-162;384-394;440-457 <YO2>  
 C:Genetics:  
 A:Map position: 4  
 C:Function:  
 A:Description: hydrolyzes 1,4-glycosidic linkages of starch, removing maltose units succ  
 C:Superfamily: beta-amylase  
 C:Keywords: glycosidase; hydrolase; monomer; polysaccharide degradation; seed; tandem re  
 F:2-535/Product: beta-amylase #status predicted <MAT>  
 F:488-531/Region: 11-residue repeats  
 F:184,378/Active site: Glu #status predicted

Query Match 0.8%; Score 8; DB 1; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811  
 DB 55 GLVEGKGP 62  
 |||||

RESULT 14  
 S71365  
 ovule development protein aintegumenta - Arabidopsis thaliana  
 N:Alternate names: AP2 domain-containing protein; protein T28119.30  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 09-Jul-2004  
 C:Accession: S71365; T06022  
 R:Klucher, K.  
 submitted to the EMBL Data Library, November 1995  
 A:Reference number: S71365  
 A:Accession: S71365  
 A:Molecule type: mRNA  
 A:Residues: 1-555 <KLU>  
 A:Cross-references: UNIPROT:Q42462; EMBL:U40256; NID:gl209098; PID:gl209099

R.; Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15484

A:Accession: T06022

A:Molecule type: DNA

A:Residues: 1-555 <REV>

A:Cross-references: EMBL:AL035709; GSPDB:GN00062; ATSP:T28119.30

A:Experimental source: cultivar Columbia; BAC clone T28119

C:Genetics:

A:Gene: ATSP:T28119.30; ANT

A:Map position: 4

A:Introns: 111/1; 289/2; 317/1; 320/1; 349/3; 391/2; 417/1

Query Match 0.8%; Score 8; DB 2; Length 555;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVLTL 12

DB 536 LKTVLTL 543

RESULT 15

JC5929

serine/arginine-rich protein-specific kinase (EC 2.-.-.-) 2 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 10-Apr-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004

C:Accession: JC5929

R.; Kuroyanagi, N.; Onogi, H.; Wakabayashi, T.; Hagiwara, M.

Biochem. Biophys. Res. Commun. 242, 357-364, 1998

A:Title: Novel SR-protein-specific kinase, SRPK2, disassembles nuclear speckles.

A:Reference number: JC5929; MUID:98113357; PMID:9446799

A:Accession: JC5929

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-681 <KOR>

A:Cross-references: UNIPROT:Q8VCD9; UNIPROT:O54781

A:Experimental source: brain

C:Comment: This enzyme regulates the disassembly of the serine/arginine-rich proteins in a

C:Keywords: transferase

Query Match 0.8%; Score 8; DB 2; Length 681;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSE 951

DB 5 SEKSSSE 12

RESULT 16

A87330

sensory box histidine kinase/response regulator [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004

C:Accession: A87330

R.; Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Esmailova, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: A87330

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-776 <STO>

A:Cross-references: UNIPROT:Q9AAB9; GB:AE005673; NID:gl3421867; PIDN:AAK22637.1; GSPDB:G

C:Genetics:

A:Gene: CC0652

Query Match 0.8%; Score 8; DB 2; Length 776;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 716 TAGKALLS 723

DB 455 TAGKALLS 462

RESULT 17

T49593

hypothetical protein B3E4.60 [imported] - Neurospora crassa

C:Species: Neurospora crassa

C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jul-2004

C:Accession: T49593

R.; Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49593

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-862 <SCH>

A:Cross-references: UNIPROT:Q9PSX6; EMBL:AL355931; GSPDB:GN00116; NCSP:B3E4.60

A:Experimental source: BAC clone B3E4; strain OR74A

C:Genetics:

A:Gene: NCSP:B3E4.60

A:Map position: 6

A:Introns: 181/2; 764/1

C:Superfamily: Saccharomyces cerevisiae probable membrane protein YOR165w

Query Match 0.8%; Score 8; DB 2; Length 862;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 493 KSFEASDD 500

DB 573 KSFEASDD 580

RESULT 18

SVBSVS

valine-tRNA ligase (EC 6.1.1.9) - Bacillus stearothermophilus

N:Alternate names: valyl-tRNA synthetase

C:Species: Bacillus stearothermophilus

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004

C:Accession: A26738

R.; Borgford, T.J.; Brand, N.J.; Gray, T.E.; Fersht, A.R.

Biochemistry 26, 2480-2486, 1987

A:Title: The valyl-tRNA synthetase from Bacillus stearothermophilus has considerable seq

A:Reference number: A26738; MUID:87271643; PMID:3300774

A:Accession: A26738

A:Molecule type: DNA

A:Residues: 1-880 <BOR>

A:Cross-references: UNIPROT:P11931; GB:M16318

C:Genetics:

A:Gene: valS

C:Superfamily: valine-tRNA ligase

C:Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis

Query Match 0.8%; Score 8; DB 1; Length 880;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 793 KRLEKELD 800

DB 820 KRLEKELD 827

RESULT 19

T30299

dynein heavy chain isotype 7A - sea urchin (Tripneustes gratilla) (fragment)

C:Species: Tripneustes gratilla

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T30299

R.; Gibbons, B.H.; Asai, D.J.; Tang, W.J.; Hays, T.S.; Gibbons, I.R.

Mol. Biol. Cell 5, 57-70, 1994

A>Title: Phylogeny and expression of axonemal and cytoplasmic dynein genes in sea urchin  
 A:Reference number: Z20814; MUID:94243035; PMID:8186465

A:Accession: T30299

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1114 <GIB>

A:Cross-references: UNIPROT:Q27811; EMBL:U03978; NID:G433384; PID:G433385; PIDN:AAA63592

C:Genetics:

A:Gene: DYH7A

C:Superfamily: dynein heavy chain, ciliary

Query Match 0.8%; Score 8; DB 2; Length 1114;

Best Local Similarity 100.0%; Pred. No. 68;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 275 DLPDLNKA 282

Db 152 DLPDLNKA 159

#### RESULT 20

JC4889

phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase (EC 3.1.3.-) - human

N:Alternate names: hps1CN5phosphatase

C:Species: Homo sapiens (man)

C>Date: 10-Sep-1996 #sequence\_revision 18-Oct-1996 #text\_change 09-Jul-2004

C:Accession: JC4889; PC4187

R:Drayner, A.L.; Peesee, X.; De Smedt, F.; Woeholtski, R.; Parker, P.; Erneux, C.

Biochem. Biophys. Res. Commun. 225, 243-249, 1996

A>Title: Cloning and expression of a human placenta inositol 1,3,4,5-tetrakisphosphate

A:Reference number: JC4889; MUID:96332436; PMID:8769125

A:Contents: Placenta

A:Accession: JC4889

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1188 <DRA>

A:Cross-references: UNIPROT:Q92656; EMBL:X98429; NID:G1495455; PIDN:CAA67071.1; PID:e249

A:Accession: PC4187

A:Molecule type: protein

A:Residues: 582-592; 668-675 <DR2>

C:Comment: This enzyme plays a function in the phosphatidylinositol 3-kinase and isositol

sequence motifs show that this enzyme interacts with various proteins in signal transdu

C:Comment: Inositol 1,3,4,5-tetrakisphosphate is also a substrate for this enzyme.

C:Keywords: phosphoric monoester hydrolase

F:5-101/Domain: SH2 homology <SH2>

F:380/Binding site: substrate (Arg) #status predicted

F:671/Active site: Cys #status predicted

Query Match 0.8%; Score 8; DB 2; Length 1188;

Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 901 SKLQLTS 908

Db 262 SKLQLTS 269

#### RESULT 21

VGBE11

149K glycoprotein - ictaluriid herpesvirus 1 (strain auburn 1)

C:Species: ictaluriid herpesvirus 1

A>Note: host Ictalurus punctatus (channel catfish)

C>Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 09-Jul-2004

C:Accession: B36791

R:Davidson, A.J.

submitted to GenBank, January 1992

A:Description: Channel catfish virus: a new type of herpesvirus.

A:Reference number: A36804

A:Accession: B36791

A:Molecule type: DNA

A:Residues: 1-1355 <DAV>

A:Cross-references: UNIPROT:Q00104; GB:M75136; NID:G331209; PIDN:AAA88149.1; PID:G331256

R:Davidson, A.J.

Virology 186, 9-14, 1992

A>Title: Channel catfish virus: a new type of herpesvirus.

A:Reference number: A39447; MUID:92087490; PMID:1727613

A:Contents: annotation

A>Note: neither amino acid nor nucleotide sequence is given

C:Genetics:

A:Gene: 46

C:Superfamily: ictaluriid herpesvirus 149K glycoprotein

C:Keywords: Glycoprotein

F:81.112,129,169,173,192,542,655,682,744,780,811,815,860,865,868,882,895,1195,1213,1225,

Query Match 0.8%; Score 8; DB 1; Length 1355;

Best Local Similarity 100.0%; Pred. No. 81;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 881 ALAVATLA 888

Db 962 ALAVATLA 969

#### RESULT 22

T08880

NMDA receptor-binding protein yotiao - human

C:Species: Homo sapiens (man)

C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: T08880

R:Lin, J.W.; Mysznyski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.

J. Neurosci. 18, 2017-2027, 1998

A>Title: Yotiao, a novel protein of neuromuscular junction and brain that interacts with

A:Reference number: Z16511; MUID:98151389; PMID:9482789

A:Accession: T08880

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1642 <LIN>

A:Cross-references: EMBL:AF026245; NID:G2623067; PIDN:AAB6384.1; PID:G2623068

C:Genetics:

A:Map position: 7q21-22

C:Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction; skeletal muscle

Query Match 0.8%; Score 8; DB 2; Length 1642;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKDKL 523

Db 357 TLELKDKL 364

#### RESULT 23

D90841

hypothetical protein ECe1700 [imported] - Escherichia coli (strain O157:H7, substrain RI

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004

C:Accession: D90841

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90841

A:Molecule type: DNA

A:Residues: 1-66 <HAY>

A:Cross-references: UNIPROT:Q8XDH8; GB:BA000007; PIDN:BA835123.1; PID:G13361165; GSPDB:G

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECe1700

Query Match 0.7%; Score 7; DB 2; Length 66;

Best Local Similarity 100.0%; Pred. No. 58;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 970 RHSAILP 976

```

Db      50 RHSAILP 56
|||||
RESULT 24
C85699
hypothetical protein Z1967 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: C85699
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85699
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-66 <STO>
A:Cross-references: UNIPROT:Q8XDH8; GB:AE005174; NID:q12514907; PIDN:AA656055.1; GSPDB:C
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1967

Query Match      0.7%; Score 7; DB 2; Length 66;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      970 RHSAILP 976
|||||
Db      50 RHSAILP 56
|||||

RESULT 25
A82086
hypothetical protein VC2357 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: A82086
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: A82086
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-77 <HE1>
A:Cross-references: UNIPROT:Q9KPL0; GB:AE004306; GB:AE003852; NID:g9656924; PIDN:AAF9550
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2357
A:Map position: 1

Query Match      0.7%; Score 7; DB 2; Length 77;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      793 KRLEKEL 799
|||||
Db      57 KRLEKEL 63
|||||

RESULT 26
A60457
Clusterin - sheep (fragments)
N:Alternin - sheep (fragments)
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 19-Oct-1995
C:Accession: A60457; S10671; S10672
R:Griswold, M.D.; Bishop, P.D.; Kim, K.H.; Ping, R.; Siiteri, J.E.; Morales, C.
Ann. N. Y. Acad. Sci. 564, 154-172, 1989

```

```

A:Title: Function of vitamin A in normal and synchronized seminiferous tubules.
A:Reference number: A60457; MUID:89372568; PMID:2672955
A:Accession: A60457
A:Molecule type: protein
A:Residues: 1-82 <GR1>
R:Tsuruta, J.K.; Wong, K.; Fritz, I.B.; Griswold, M.D.
Biochem. J. 268, 571-578, 1990
A:Title: Structural analysis of sulphated glycoprotein 2 from amino acid sequence. Relat
A:Reference number: S10671; MUID:90303244; PMID:2363694
A:Accession: S10671
A:Molecule type: protein
A:Residues: 4-26; 41-49 <TSU>
A:Accession: S10672
A:Molecule type: protein
A:Residues: 62-82 <TS2>
C:Superfamily: clusterin
C:Keywords: glycoprotein; testis

Query Match      0.7%; Score 7; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      603 NALKEVL 609
|||||
Db      26 NALKEVL 32
|||||

RESULT 27
H69318
conserved hypothetical protein AF0552 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: H69318
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: H69318
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-88 <KLE>
A:Cross-references: UNIPROT:O29699; GB:AE001066; GB:AE000782; NID:g2689389; PIDN:AA89068
C:Superfamily: probable sulfur carrier protein sir0821
F:88/Modified site: 1-thioglycine (Gly) #status predicted

Query Match      0.7%; Score 7; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      687 DLEKILK 693
|||||
Db      39 DLEKILK 45
|||||

RESULT 28
D69873
hypothetical protein ylan - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: D69873
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertel
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

```

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Aubors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Siroz, A.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, M.; Yashiki, H.; Yoshikawa, H.F.; Zumstein, B.; Yoshikawa, H.; Danchin, A.  
A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A'Reference number: A69580; MUID:98044033; PMID:9384377  
A'Accession: D69873  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A'Molecule type: DNA  
A'Residuals: 1-93 <DNA>  
A'CROSS-references: UNIPROT:007638; GB:Z99111; GB:AL009126; NID:G2633699; PIDN:CAB13357  
A'Experimental source: strain 168  
C'Genetics:  
A'Gene: ylan

Query Match 0.7%; Score 7; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 689 EKILKLI 695  
|||||  
Db 22 EKILKLI 28

RESULT 29  
C64431  
conserved hypothetical protein MJ1052 - *Methanococcus jannaschii*  
C'Species: *Methanococcus jannaschii*  
C'Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
C'Accession: C64431  
R'Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Reon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A'Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.  
A>Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*  
A'Reference number: A64300; MUID:96337999; PMID:8688087  
A'Accession: C64431  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A'Molecule type: DNA  
A'Residuals: 1-100 <BUL>  
A'CROSS-references: UNIPROT:Q58452; GB:U67548; GB:L77117; NID:G2826361; PIDN:AAB99055.1;  
C'Genetics:  
A'Map position: REV989804-989502  
A'Start codon: GTG  
C'Superfamily: *Methanococcus jannaschii* conserved hypothetical protein MJ1052

Query Match 0.7%; Score 7; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 280 LKAIGEL 286  
|||||  
Db 94 LKAIGEL 100

RESULT 30  
TRVKGI  
terminal region recognition factor 1 - Yeast (*Kluyveromyces marxianus* var. *lactis*) plasmid  
N'Alternate names: DNA-binding protein TRF1  
C'Species: *Kluyveromyces marxianus* var. *lactis*, *Candida spharica*  
A'Note: plasmid pGK12, K2  
C'Date: 30-Sep-1989 #sequence\_revision 09-May-1997 #text\_change 09-Jul-2004  
C'Accession: S00969; S10336; S17251; A41628  
R'Tommasino, M.; Ricci, S.; Galeotti, C.L.  
Nucleic Acids Res. 16, 5863-5878, 1988  
A>Title: Genome organization of the killer plasmid pGK12 from *Kluyveromyces lactis*.  
A'Reference number: S00959; MUID:88289339; PMID:3041369  
A'Accession: S00968  
A'Molecule type: DNA  
A'Residuals: 1-103 <TOM>  
A'CROSS-references: UNIPROT:P05476; EMBL:X07776; NID:G2868; PIDN:CAA30611.1; PID:G2878

R;Wilson, D.W.; Meacock, P.A.  
Nucleic Acids Res. 16, 8097-8112, 1988  
A>Title: Extracellular gene expression in yeast: evidence for a plasmid-encoded RNA polymerase.  
A'Reference number: S10336; MUID:98335549; PMID:3138657  
A'Accession: S10336  
A>Status: nucleic acid sequence not shown; translation not shown  
A'Molecule type: DNA  
A'Residuals: 1-103 <WIL>  
A'CROSS-references: EMBL:X07946; NID:G2883; PIDN:CAA30767.1; PID:G2884  
A'Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1988  
R;McNeel, D.G.; Tamanoi, F.  
Proc. Natl. Acad. Sci. U.S.A. 88, 11398-11402, 1991  
A>Title: Terminal region recognition factor 1, a DNA-binding protein recognizing the invin-1 sequence.  
A'Reference number: A41628; MUID:92107957; PMID:1763054  
A'Contents: annotation; purification and activity after expression in *E. coli*  
R;Tommasino, M.  
Yeast 7, 245-252, 1991  
A>Title: Killer system of *Kluyveromyces lactis*: the open reading frame 10 of the pGK12 p  
A'Reference number: S17251; MUID:91353078; PMID:1882549  
A'Accession: S17251  
A'Molecule type: DNA  
A'Residuals: 1-71; 83-99 <TO2>  
A'CROSS-references: PIDN:AAB19610.1; PID:G234312; PIDN:AAB19611.1; PID:G234313  
C'Genetics:  
A'Genome: plasmid  
C'Function:  
A>Description: DNA binding; may play a role in the protein-primed initiation of the repl  
C'Superfamily: DNA-binding protein TRF1  
C'Keywords: DNA binding; DNA replication; plasmid replication

Query Match 0.7%; Score 7; DB 1; Length 103;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 999 LITAICK 1005  
|||||  
Db 9 LITAICK 15

RESULT 31  
GEMSM1  
matrix Gla protein precursor - mouse  
N'Alternate names: MGP  
C'Species: *Mus musculus* (house mouse)  
C'Date: 07-Sep-1990 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C'Accession: JQ0455; I55484  
R;Ikeda, T.  
submitted to JIPID, May 1990  
A'Reference number: JQ0455  
A'Accession: JQ0455  
A'Molecule type: mRNA  
A'Residuals: 1-104 <IKE>  
A'CROSS-references: UNIPROT:P19788; DBJ:D00613; NID:G220478; PIDN:BAA00488.1; PID:G2204  
A'Experimental source: strain C57BL/6, cell type osteoblast  
R;Luo, G.; D'Souza, R.; Hogue, D.; Karsenty, G.  
J. Bone Miner. Res. 10, 325-334, 1995  
A>Title: The matrix Gla protein gene is a marker of the chondrogenesis cell lineage during  
A'Reference number: I55484; MUID:95274434; PMID:7754814  
A'Accession: I55484  
A>Status: translation not shown; translated from GB/EMBL/DBJ  
A'Molecule type: DNA  
A'Residuals: 1-24, 1', 26-104 <LUO>  
A'CROSS-references: GB:S77350; NID:G998473  
A'Experimental source: 129 SvEv  
A'Note: the source is designated as mouse and classified as *Mus sp.* by NCBI  
C'Comment: This vitamin K-dependent calcium binding protein is secreted by most known ce  
than osteocalcin.  
C'Genetics:  
A'Introns: 21/1; 32/1; 57/2  
C'Superfamily: osteocalcin  
C'Keywords: bone; calcium binding; carboxyglutamic acid; cartilage; phosphoprotein; vita  
P;1-19/Domain: signal sequence #status predicted <SIG>  
P;20-104/Product: matrix Gla protein #status predicted <MAT>

F;21,60,67,71/Modified site: gamma-carboxyglutamic acid (Glu) #status predicted  
F;22,25,28/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted  
F;73-79/disulfide bonds: #status predicted

Query Match 0.7%; Score 7; DB 1; Length 104;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 881 ALAVATL 887  
|||||

DB 12 ALAVATL 18

#### RESULT 32

T05603  
H+-exporting ATPase (EC 3.6.3.6) chain G, vacuolar - Arabidopsis thaliana  
N:Alternate names: protein F9D16.180  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T05603  
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.P.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15419  
A:Accession: T05603  
A:Molecule type: DNA  
A:Residues: 1-106 <BEV>  
A:Cross-references: UNIPROT:O82629; EMBL:AL035394  
A:Experimental source: Cultivar Columbia; BAC clone F9D16  
C:Genetics:  
A:Gene: vag2  
A:Map position: 4  
A:Introns: 26/1; 59/3  
A:Note: F9D16.180  
C:Superfamily: yeast H+-transporting ATPase 13K chain  
C:Keywords: ATP; hydrogen ion transport; hydrolase

Query Match 0.7%; Score 7; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 790 ANVKRL 796  
|||||

DB 66 ANVKRL 72

#### RESULT 33

T51826  
Hypothetical protein vag2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 09-Jul-2004  
C:Accession: T51826  
R:Rouquie, D.; Tournaire-Roux, C.; Szponarski, W.; Rossignol, M.; Dumas, P.  
submitted to the EMBL Data Library, May 1998  
A:Description: Cloning and expression of G subunits of vacuolar-type ATPase from plants.  
A:Reference number: Z25473  
A:Accession: T51826  
A:Molecule type: mRNA  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Residues: 1-106 <ROU>  
A:Cross-references: UNIPROT:O82629; EMBL:AJ005902; PIDN:CAA06759.1  
C:Genetics:  
A:Gene: vag2  
C:Superfamily: yeast H+-transporting ATPase 13K chain

Query Match 0.7%; Score 7; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 790 ANVKRL 796  
|||||

DB 66 ANVKRL 72

#### RESULT 34

T51825  
H+-exporting ATPase (EC 3.6.3.6) chain G, vacuolar [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 09-Jul-2004  
C:Accession: T51825  
R:Rouquie, D.; Tournaire-Roux, C.; Szponarski, W.; Rossignol, M.; Dumas, P.  
submitted to the EMBL Data Library, May 1998  
A:Description: Cloning and expression of G subunits of vacuolar-type ATPase from plants  
A:Reference number: Z25473  
A:Accession: T51825  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-110 <ROU>  
A:Cross-references: UNIPROT:O82628; EMBL:AJ005901; PIDN:CAA06758.1  
A:Experimental source: Cultivar Columbia; leaves  
C:Genetics:  
A:Gene: vag1  
C:Superfamily: yeast H+-transporting ATPase 13K chain  
C:Keywords: ATP; hydrogen ion transport; hydrolase

Query Match 0.7%; Score 7; DB 2; Length 110;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 790 ANVKRL 796  
|||||

DB 70 ANVKRL 76

#### RESULT 35

E84325  
Hypothetical protein Vng1734h [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: E84325  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, E.; Leitchauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablun, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: E84325  
A:Molecule type: preliminary  
A:Status: preliminary  
A:Residues: 1-111 <STO>  
A:Cross-references: UNIPROT:Q9HPAL; GB:AE004437; NID:g10581194; PIDN:AAG19969.1; GSPDB:G  
C:Genetics:  
A:Gene: VNG1734H

Query Match 0.7%; Score 7; DB 2; Length 111;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 882 LAVATLA 888  
|||||

DB 44 LAVATLA 50

#### RESULT 36

E95881  
Hypothetical protein [imported] - Sinorhizobium meliloti (strain 1021) magaplasamid pSymE  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: E95881  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhxwester, J.; Chain, P.; Vorholter, F.J.; Hernan  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1,683-kb pSymB megaplasamid from the N2-fixing endo  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: E95881  
A:Status: preliminary  
A:Molecule type: DNA

QY 790 ANVKRL 796  
|||||

DB 66 ANVKRL 72



A;Residues: 1-118 <KUR>  
A;Cross-references: UNIPROT:Q92NM2; GB:AL591985; PIDN:CAC48717.1; PID:gl5140190; GSPDB:G  
A;Experimental source: strain 1021, megaplasmid pSymB  
R;Galibart, F.; Finan, T.M.; Long, S.R.; Fuhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler,  
pala, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
heault, P.; Vandenbot, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A;Reference number: A96039; MUID:21368234; PMID:11474104  
C;Contents: annotation  
C;Genetics:  
A;Gene: SMB20331  
A;Genome: plasmid

Query Match 0.7%; Score 7; DB 2; Length 118;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 881 ALAVATL 887  
|||||  
Db 51 ALAVATL 57

RESULT 37  
B96647  
hypothetical protein F19K23.2 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: B96647  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: B96647  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-121 <STO>  
A;Cross-references: UNIPROT:Q04574; GB:AE005173; NID:G2160134; PIDN:AAB60756.1; GSPDB:GN  
C;Genetics:  
A;Gene: F19K23.2  
A;Map position: 1

Query Match 0.7%; Score 7; DB 2; Length 121;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 825 KTPLELP 831  
|||||  
Db 46 KTPLELP 52

RESULT 38  
QQVZEL  
Flu protein - vaccinia virus  
C;Species: vaccinia virus  
C;Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 09-Jul-2004  
C;Accession: B26216; L23092  
R;Plucieniczak, A.; Schroeder, E.; Zettlmeisel, G.; Streeck, R.E.  
Nucleic Acids Res. 13, 985-998, 1985  
A;Title: Nucleotide sequence of a cluster of early and late genes in a conserved segment  
A;Reference number: A23092; MUID:85215527; PMID:2987815  
A;Accession: B26216  
A;Molecule type: DNA  
A;Residues: 1-124 <PLU>

A;Cross-references: UNIPROT:P07618; GB:X01978; GB:J02424; GB:J02425; GB:K02376; GB:ML521  
A;Experimental source: strain WR  
C;Comment: It is unknown whether this protein is synthesized before or after viral DNA r  
C;Superfamily: vaccinia virus F11 protein

Query Match 0.7%; Score 7; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 LLPASLK 615  
|||||  
Db 52 LLPASLK 58

RESULT 39  
S28695  
hypothetical protein 13 - Agrobacterium tumefaciens plasmid pTi15955  
C;Species: Agrobacterium tumefaciens  
C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Jul-2004  
C;Accession: S28695  
R;Barker, R.F.; Idler, K.B.; Thompson, D.V.; Kemp, J.D.  
Plant Mol. Biol. 2, 335-350, 1983  
A;Title: Nucleotide sequence of the T-DNA region from the Agrobacterium tumefaciens octo  
A;Reference number: S28683  
A;Accession: S28695  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-124 <BAR>  
A;Cross-references: UNIPROT:Q44392; EMBL:X00493; NID:G39062; PIDN:CAA25175.1; PID:G39075  
C;Genetics:  
A;Genome: plasmid

Query Match 0.7%; Score 7; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 970 RHSAILP 976  
|||||  
Db 5 RHSAILP 11

RESULT 40  
H71137  
hypothetical protein PH0868 - Pyrococcus horikoshii  
C;Species: Pyrococcus horikoshii  
C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 09-Jul-2004  
C;Accession: H71137  
R;Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
A;Reference number: A71000; MUID:98344137; PMID:9679194  
A;Accession: H71137  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-124 <KAW>  
A;Cross-references: UNIPROT:O58598; GB:AF000003; NID:G3236130; PIDN:BAA29962.1; PID:G325  
A;Experimental source: strain OT3  
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C;Genetics:  
A;Gene: PH0868

Query Match 0.7%; Score 7; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 994 VALLSLI 1000  
|||||  
Db 18 VALLSLI 24

RESULT 41  
C95270

hypothetical protein Sma0128 [imported] - Sinorhizobium meliloti (strain 1021) megaplasmid  
 C:Species: Sinorhizobium meliloti  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 09-Jul-2004  
 C:Accession: C95270  
 R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bova, A.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C. Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti A:Residues: 1-133 <JOH>  
 A:Cross-references: A95262; MUID:21396509; PMID:11481432  
 A:Accession: C95270  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-133 <KUR>  
 A:Cross-references: UNIPROT:Q930X7; GB:AE006469; PIDN:AAK64725.1; PID:g14523128; GSPDB:G A:Experimental source: strain 1021, megaplasmid pSymA  
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T. Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, Hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K. A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: Sma0128  
 A:Genome: plasmid

Query Match 0.7%; Score 7; DB 2; Length 126;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 929 AAAREAQ 935  
 Db 79 AAAREAQ 85

## RESULT 42

H81371  
 hypothetical protein Cj0971 [imported] - Campylobacter jejuni (strain NCTC 11168)  
 C:Species: Campylobacter jejuni  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
 C:Accession: H81371  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling, C.W.; Quail, M.; Ralston, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell, Nature 403, 665-668, 2000  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp A:Reference number: A81250; MUID:20150912; PMID:10688204  
 A:Accession: H81371  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-130 <PAR>  
 A:Cross-references: UNIPROT:Q9PNW7; GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CB7322 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: Cj0971  
 C:Superfamily: Campylobacter jejuni hypothetical protein Cj0971

Query Match 0.7%; Score 7; DB 2; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 NALKEVL 609  
 Db 99 NALKEVL 105

## RESULT 43

H42513  
 J5L 15.2K protein - vaccinia virus (strain Ankara, Copenhagen)  
 C:Species: vaccinia virus  
 A:Note: host Homo sapiens (man)  
 C:Date: 09-Nov-1990 #sequence\_revision 09-Nov-1990 #text\_change 09-Jul-2004

C:Accession: H42513; T37365  
 R:Johnson, G.P. submitted to GenBank, June 1990  
 A:Reference number: A31172  
 A:Accession: H42513  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-133 <JOH>

A:Cross-references: UNIPROT:P21083  
 R:Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dorner, F. submitted to the EMBL Data Library, March 1997  
 A:Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) strain A:Reference number: Z20877

A:Accession: T37365  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-133 <ANT>  
 A:Cross-references: EMBL:U94848; PIDN:AB964445.1  
 A:Experimental source: strain Ankara  
 C:Genetics:  
 A:Note: MVA089L  
 C:Superfamily: vaccinia virus F11 protein

Query Match 0.7%; Score 7; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 LLPASLK 615  
 Db 52 LLPASLK 58

## RESULT 44

T28520  
 hypothetical protein L5L - variola major virus  
 C:Species: variola major virus  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
 C:Accession: T28520  
 R:Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Aubin, Nature 366, 748-751, 1993  
 A:Title: Potential virulence determinants in terminal regions of variola smallpox virus A:Reference number: Z20488; MUID:94088747; PMID:8264798  
 A:Accession: T28520  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-133 <MAS>  
 A:Cross-references: UNIPROT:Q85384; EMBL:L22579; NID:G623595; PIDN:AAA60830.1; PID:G43930 A:Experimental source: strain Bangladesh-1975  
 C:Superfamily: vaccinia virus F11 protein

Query Match 0.7%; Score 7; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 LLPASLK 615  
 Db 52 LLPASLK 58

## RESULT 45

H72160  
 M5L protein - variola minor virus (strain Garcia-1966)  
 C:Species: variola minor virus  
 C:Date: 24-Nov-1999 #sequence\_revision 24-Nov-1999 #text\_change 09-Jul-2004  
 C:Accession: H72160  
 R:Shchelkunov, S.N.; Totmenin, A.V.; Gutorov, V.V.; Safronov, P.F.; Massung, R.F.; Lopar submitted to GenBank, March 1998  
 A:Description: Analysis of the complete coding sequence of DNA of alastrim variola minor A:Reference number: A72150

A:Accession: H72160  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-133 <SHC>

A;Cross-references: UNIPROT:P33055; GB:Y16780; NID:G5830555; PIDN:CAB54682.1; PID:G58306  
 A;Experimental source: strain Garcia-1966  
 C;Genetics:  
 A;Gene: M5L  
 C;Superfamily: vaccinia virus F11 protein

Query Match 0.7%; Score 7; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 LLPASLK 615  
 |||||  
 Db 52 LLPASLK 58

RESULT 46  
 S33096  
 J5L protein - variola virus  
 N;Alternate names: L5L protein  
 C;Species: variola virus  
 C;Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 09-Jul-2004  
 C;Accession: S33096; G36845  
 R;Shchelkunov, S.N.; Blinov, V.M.; Totmenin, A.V.; Marennikova, S.S.; Kolykhalov, A.A.;  
 dshapardze, O.G.; Sandakhchiev, L.S.  
 Virus Res. 27, 25-35, 1993  
 A;Title: Nucleotide sequence analysis of variola virus HindIII M, L, I genome fragments.  
 A;Reference number: S33069; MUID:93190624; PMID:8383392  
 A;Accession: S33096  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-133 <SHC>  
 A;Cross-references: UNIPROT:P33055; EMBL:X67119; NID:G62330; PIDN:CAA47581.1; PID:G62358  
 A;Experimental source: strain India-1967, isolate Ind3  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992  
 R;Blinov, V.M.  
 submitted to GenBank, November 1992  
 A;Reference number: A36859  
 A;Accession: G36845  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-133 <BLI>  
 A;Cross-references: GB:X69198; NID:G456758; PIDN:CAA49023.1; PID:G297262  
 A;Experimental source: strain India-1967, sep. major, isolate Ind3  
 C;Superfamily: vaccinia virus F11 protein

Query Match 0.7%; Score 7; DB 2; Length 133;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 LLPASLK 615  
 |||||  
 Db 52 LLPASLK 58

RESULT 47  
 WMM516  
 submandibular gland 16.5K protein - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 09-Jul-2004  
 C;Accession: A03298  
 R;Windas, J.D.; Mullins, J.J.; Beecroft, L.J.; George, H.; Meacock, P.A.; Williams, B.R.  
 Nucleic Acids Res. 12, 1361-1376, 1984  
 A;Title: Molecular cloning of cDNAs from androgen-independent mRNA species of DBA/2 mouse  
 A;Reference number: A93503; MUID:84144035; PMID:6546617  
 A;Accession: A03298  
 A;Molecule type: mRNA  
 A;Residues: 1-138 <WIN>  
 A;Cross-references: UNIPROT:P02815; GB:X00349; NID:G51367; PIDN:CAA95098.1; PID:G51368  
 C;Comment: this protein contains a hydrophobic amino-terminal sequence that is similar to  
 C;Superfamily: submandibular gland 16.5K protein  
 A;Keywords: glycoprotein; submandibular gland  
 F;25,72,89,94/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 0.7%; Score 7; DB 1; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 KSGVEKL 157  
 |||||  
 Db 125 KSGVEKL 131

RESULT 48  
 A30305  
 submandibular gland protein (spot 1) precursor - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 22-Nov-1989 #sequence\_revision 03-Aug-1992 #text\_change 09-Jul-2004  
 C;Accession: A30305  
 R;Dickinson, D.P.; Mirels, L.; Tabak, L.A.; Gross, K.W.  
 Mol. Biol. Evol. 6, 80-102, 1989  
 A;Title: Rapid evolution of variants in a rodent multigene family encoding salivary protein  
 A;Reference number: A30305; MUID:89158788; PMID:2921944  
 A;Accession: A30305  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-138 <DIC>  
 A;Cross-references: UNIPROT:P02815; GB:M33974; NID:G201044; PIDN:AAA40132.1; PID:G201045  
 A;Note: the authors translated the codon AAC for residue 104 as Gln  
 C;Superfamily: submandibular gland 16.5K protein

Query Match 0.7%; Score 7; DB 2; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 KSGVEKL 157  
 |||||  
 Db 125 KSGVEKL 131

RESULT 49  
 T35860  
 hypothetical protein SC9B1.15c - Streptomyces coelicolor  
 C;Species: Streptomyces coelicolor  
 C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004  
 C;Accession: T35860  
 R;Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.  
 submitted to the EMBL Data Library, April 1999  
 A;Reference number: Z21591  
 A;Accession: T35860  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-139 <SAU>  
 A;Cross-references: UNIPROT:Q9X835; EMBL:AL049727; PIDN:CAB41561.1; GSPDB:GN000070; SCOD  
 A;Experimental source: strain A3(2)  
 C;Genetics:  
 A;Gene: SCODB:SC9B1.15c  
 C;Superfamily: Streptomyces coelicolor hypothetical protein SC1C2.20c

Query Match 0.7%; Score 7; DB 2; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 801 LLTGLVE 807  
 |||||  
 Db 15 LLTGLVE 21

RESULT 50  
 A82517  
 ribonuclease P XP2781 [imported] - Xylella fastidiosa (strain 9a5c)  
 C;Species: Xylella fastidiosa  
 C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 16-Aug-2004  
 C;Accession: A82517  
 R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 Nature 406, 151-157, 2000  
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: For a complete list of authors see reference number A59328 below  
A:Accession: A82517  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-142 <SIM>  
A:Cross-references: UNIPROT:Q9P9U0; GB:AE004083; GB:AE003849; NID:99108022; PIDN:AAF8556  
A:Experimental source: strain 9a5c  
R:Simpsom, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranee, E.E.; Laig  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
P.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XP2781  
C:Superfamily: bacterial ribonuclease P, protein component  
Query Match 0.7%; Score 7; DB 2; Length 142;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 993 SVALLSL 999  
Db 109 SVALLSL 115  
RESULT 51  
JC6173  
single-stranded DNA-binding protein 1 precursor, mitochondrial - African clawed frog  
N:Alternate names: helix-desstabilizing protein; single-stranded DNA-binding protein s  
C:Species: Xenopus laevis (African clawed frog)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: JC6173; S22300; S20262; S0116  
R:Champagne, A.M.; Dufresne, C.; Viney, L.; Gueride, M.  
Gene 184, 65-71, 1997  
A>Title: Cloning, sequencing and expression of the two genes encoding the mitochondrial  
A:Reference number: JC6173; MUID:97169147; PMID:9016954  
A:Contents: oocyte  
A:Accession: JC6173  
A:Molecule type: mRNA  
A:Residues: 1-146 <CHA>  
A:Cross-references: UNIPROT:P09380; EMBL:X83673; NID:9620127; PIDN:CAA58647.1; PID:91890  
R:Tiranti, V.; Barat-Gueride, M.; Bijl, J.; Didonato, S.; Zeviani, M.  
Nucleic Acids Res. 19, 4291, 1991  
A>Title: A full-length cDNA encoding a mitochondrial DNA-specific single-stranded DNA bi  
A:Reference number: S22300; MUID:91334145; PMID:1870981  
A:Accession: S22300  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-146 <TIR>  
A:Cross-references: EMBL:X59285; NID:964898; PIDN:CAA41376.1; PID:964899  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1991  
R:Ghrir, R.; Lecaer, J.P.; Dufrene, C.; Gueride, M.  
Arch. Biochem. Biophys. 291, 395-400, 1991  
A>Title: Primary structure of the two variants of Xenopus laevis mtSSB, a mitochondrial  
A:Reference number: S20262; MUID:92061073; PMID:1952953  
A:Accession: S20262  
A:Molecule type: protein  
A:Residues: 18-142 <GHR>  
R:Mahongou, C.; Ghrir, R.; Lecaer, J.P.; Mignotte, B.; Barat-Gueride, M.  
FEBS Lett. 235, 267-270, 1988  
A>Title: The amino-terminal sequence of the Xenopus laevis mitochondrial SSB is homologo  
A:Reference number: S01116; MUID:88296837; PMID:3042458  
A:Accession: S01116

A:Molecule type: protein  
A:Residues: 18-26, 'E', 28-41 <MAH>  
A:Comment: This protein is essential for replication, repair of recombination. It is co  
d, and it modulates the level of replication and transcription mediated by mitochondrial  
C:Comment: This protein binds preferentially to single-stranded DNA.  
C:Genetics:  
A:Gene: ssb2  
A:Genome: nuclear  
A:Introns: 8/3; 27/1; 74/1; 104/2; 134/1  
C:Superfamily: single-strand binding protein; single-stranded DNA-binding protein homol  
C:Keywords: DNA replication; mitochondrion; oocyte; single-stranded DNA binding; tetram  
F1-17/Domain: transit peptide (mitochondrion) #status predicted <INP>  
F18-146/Product: single-stranded DNA-binding protein 1 #status predicted <MAT>  
F.44-129/Domain: single-stranded DNA-binding protein homology <SSD>  
Query Match 0.7%; Score 7; DB 1; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 771 LERSINK 777  
Db 24 LERSINK 30  
RESULT 52  
HBHY  
hemoglobin beta chain - golden hamster  
C:Species: Mesocricetus auratus (golden hamster)  
C>Date: 17-Dec-1982 #sequence\_revision 17-Dec-1982 #text\_change 12-Jul-2004  
C:Accession: A02410  
R:Braunitzer, G.; Schrank, B.; Stangl, A.; Wiesner, H.  
J. Chem. Soc. Pak. 2, 1-7, 1980  
A>Title: Respiration at high altitudes, phosphate-protein-interaction: sequence of the  
A:Reference number: A92758  
A:Accession: A02410  
A:Molecule type: protein  
A:Residues: 1-146 <BRA>  
A:Cross-references: UNIPROT:P02094  
C:Superfamily: globin; Globin homology  
C:Keywords: blood; chromoprotein; erythrocyte; heme; iron; metalloprotein; oxygen carri  
F13-146/Domain: globin homology <GHS>  
F163/Binding site: oxygen (His) (distal axial ligand) #status predicted  
F.92/Binding site: heme iron (His) (proximal axial ligand) #status predicted  
Query Match 0.7%; Score 7; DB 1; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 755 AEKALVT 761  
Db 6 AEKALVT 12  
RESULT 53  
S22336  
hemoglobin beta chain, major - hamster  
C:Species: Cricetinae gen. sp. (hamster)  
C>Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 12-Jul-2004  
C:Accession: S22336; S36682  
R:Lee, K.M.; Subar, M.; Li, H.; Boussios, T.  
Biochim. Biophys. Acta 1130, 343-344, 1992  
A>Title: Cloning of two adult hamster globin cDNAs (alpha and beta(major)).  
A:Reference number: S22335; MUID:92223120; PMID:1562610  
A:Accession: S22336  
A:Molecule type: mRNA  
A:Residues: 1-147 <LEE>  
A:Cross-references: EMBL:X57030  
R:Boussios, T.  
submitted to the EMBL Data Library, December 1990  
A:Reference number: S36682  
A:Accession: S36682  
A:Molecule type: mRNA  
A:Residues: 1-115, 'I', 117-147 <BOU>

A;Cross-references: EMBL:X57030; NID:g49422; PIDN:CAA40346.1; PID:g49423  
 C;Superfamily: globin; globin homology  
 C;Keywords: blood; chromoprotein; erythrocyte; heme; iron; metalloprotein; oxygen carrier  
 F;4-147/Domain: globin homology <GUB>  
 F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted  
 F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 0.7%; Score 7; DB 2; Length 147;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 755 AEKALVT 761  
 |||||  
 Db 7 AEKALVT 13

RESULT 54  
 E83057  
 hypothetical protein PA4697 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C;Species: Pseudomonas aeruginosa  
 C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
 C;Accession: E83057  
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Lim,  
 -; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A;Reference number: A82950; MUID:20437337; PMID:10984043  
 A;Accession: E83057  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-150 <STO>  
 A;Cross-references: UNIPROT:Q9HV99; GB:AB004884; GB:AE004091; NID:g9950954; PIDN:AAG0808  
 A;Experimental source: strain PA01  
 C;Genetics:  
 A;Gene: PA4697

Query Match 0.7%; Score 7; DB 2; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 752 IAKAEKA 758  
 |||||  
 Db 138 IAKAEKA 144

RESULT 55  
 S36108  
 superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - guinea pig  
 C;Species: Cavia porcellus (Guinea pig)  
 C;Date: 19-Mar-1997 #sequence\_revision 09-Aug-1997 #text\_change 09-Jul-2004  
 C;Accession: S36108; S65796  
 R;Wolf, B.; Reinecke, K.; Aumann, K.D.; Brigelius-Flohe, R.; Flohe, L.  
 Biol. Chem. Hoppe-Seyler 374, 641-649, 1993  
 A;Title: Taxonomical classification of the guinea pig based on its Cu/Zn superoxide dis  
 A;Reference number: S36108; MUID:94059414; PMID:8240718  
 A;Accession: S36108  
 A;Molecule type: protein  
 A;Residues: 1-101; IL', 104-152 <WOL>  
 A;Cross-references: UNIPROT:P33431  
 R;Yuan, H.T.; Bingle, C.D.; Kelly, P.J.  
 Biochim. Biophys. Acta 1305, 163-171, 1996  
 A;Title: Differential patterns of antioxidant enzyme mRNA expression in guinea pig lung  
 A;Reference number: S65793; MUID:96180320; PMID:8597602  
 A;Accession: S65796  
 A;Molecule type: mRNA  
 A;Residues: 6-145 <YUA>  
 A;Cross-references: EMBL:U39844  
 C;Function:  
 A;Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen  
 C;Superfamily: superoxide dismutase [Cu-Zn]  
 C;Keywords: copper; metalloprotein; oxidoreductase; zinc  
 F;45,47,62,119/Binding site: copper (His) #status predicted

F;56-145/Disulfide bonds: #status predicted  
 F;62,70,79,82/Binding site: zinc (His, His, His, Asp) #status predicted  
 F;142/Active site: Arg #status predicted

Query Match 0.7%; Score 7; DB 2; Length 152;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 803 TGLVEGK 809  
 |||||  
 Db 35 TGLVEGK 41

RESULT 56  
 T48772  
 hypothetical protein I3E11.200 [imported] - Neurospora crassa  
 C;Species: Neurospora crassa  
 C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 19-May-2000  
 C;Accession: T48772  
 R;Schulze, U.; Algen, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
 submitted to the Protein Sequence Database, April 2000  
 A;Reference number: Z24541  
 A;Accession: T48772  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-152 <SCH>  
 A;Cross-references: EMBL:ALJ353820; GSPDB:GN00112; NCSP:I3E11.200  
 A;Experimental source: cosmid contig I3E11; strain 74  
 C;Genetics:  
 A;Gene: NCSP:I3E11.200  
 A;Map position: 2  
 C;Superfamily: Neurospora crassa hypothetical protein I3E11.200

Query Match 0.7%; Score 7; DB 2; Length 152;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 440 KOHNGVT 446  
 |||||  
 Db 66 KOHNGVT 72

RESULT 57  
 E82382  
 transcription regulator AsnC family VCA1068 [imported] - Vibrio cholerae (strain N16961  
 C;Species: Vibrio cholerae  
 C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
 C;Accession: E82382  
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P.  
 L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A;Reference number: A82035; MUID:20406833; PMID:10952301  
 A;Accession: E82382  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-160 <HEI>  
 A;Cross-references: UNIPROT:Q9KKM4; GB:AB004432; GB:AE003853; NID:g9658509; PIDN:AAF9696  
 A;Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C;Genetics:  
 A;Gene: VCA1068  
 A;Map position: 2  
 C;Superfamily: regulatory protein asnC

Query Match 0.7%; Score 7; DB 2; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 792 VKRLEKE 798  
 |||||  
 Db 40 VKRLEKE 46

```

RESULT 58
I70177
dynein-like protein 7 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 09-Jul-2004
C:Accession: I70177
R:Tanaka, Y.; Zhang, Z.; Hirokawa, N.
J. Cell Sci. 108, 1883-1893, 1995
A:Title: Identification and molecular evolution of new dynein-like protein sequences in
A:Reference number: 155515; MUID:95386588; PMID:7657712
A:Accession: I70177
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-161 <RES>
A:Cross-references: UNIPROT:Q63170; GB:D26498; NID:g871910; PIDN:BAA05506.1; PID:g871911
C:Genetics:
A:Gene: DLP7

Query Match 0.7%; Score 7; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 LPDNLKA 282
DB 105 LPDNLKA 111

RESULT 59
S56905
Hypothetical protein YJL124c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein J0714
C:Species: Saccharomyces cerevisiae
C>Date: 05-May-1995 #sequence_revision 08-Sep-1995 #text_change 09-Jul-2004
C:Accession: S56905
R:Cziepluch, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56891
A:Accession: S56905
A:Molecule type: DNA
A:Residues: 1-172 <ZNI>
A:Cross-references: UNIPROT:P47017; EMBL:Z49399; NID:g1008319; PIDN:CAA89419.1; PID:g100
C:Genetics:
A:Gene: SGD:LSM1; MIPS:YJL124C
A:Cross-references: SGD:S0003660
A:Map position: 10L
C:Superfamily: hypothetical protein YJL124c

Query Match 0.7%; Score 7; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 952 SANSKDR 958
DB 2 SANSKDR 8

RESULT 60
T45501
Hypothetical protein ycid [imported] - Escherichia coli (fragment)
C:Species: Escherichia coli
C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C:Accession: T45501
R:Milman, R.
submitted to the EMBL Data Library, April 1995
A:Description: Recombinational exchange among clonal populations.
A:Reference number: Z22991
A:Accession: T45501
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-173 <ML>
A:Cross-references: UNIPROT:P21364; EMBL:U24203; PIDN:AAB60129.1
A:Experimental source: strain ECOR 52

```

```

C:Genetics:
A>Note: ycid

Query Match 0.7%; Score 7; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 881 ALAVATL 887
DB 8 ALAVATL 14

RESULT 61
T02557
Hypothetical protein At2g32580 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein T26B15.14
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 09-Jul-2004
C:Accession: T02557; H84734
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC T26B15 genomic sequence.
A:Reference number: Z14678
A:Accession: T02557
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-175 <ROU>
A:Cross-references: UNIPROT:O80895; EMBL:AC004681; NID:g3298532; PID:g3298546
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: H84734
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-175 <STO>
A:Cross-references: GB:AE002093; NID:g3298546; PIDN:AAC25940.1; GSPDB:GN00139
C:Genetics:
A:Gene: T26B15.14; At2g32580
A:Map position: 2
A:Introns: 61/1

Query Match 0.7%; Score 7; DB 2; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 754 KAEKALV 760
DB 133 KAEKALV 139

RESULT 62
H86805
Hypothetical protein yogM [imported] - Lactococcus lactis subsp. lactis (strain ILL403)
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: H86805
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s8
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: H86805
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-180 <STO>
A:Cross-references: UNIPROT:Q9CFL9; GB:AE005176; PID:g12724438; PIDN:AAK05546.1; GSPDB:G
A:Experimental source: strain ILL403
C:Genetics:
A:Gene: yogM

```



```
Query Match          0.7%; Score 7; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 769 QLLERSI 775
Db 36 QLLERSI 42

RESULT 63
T32619
hypothetical protein F42A6.6 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C/Accession: T32619
R/Du, Z.; Scheet, P.; Andrews, S.
submitted to the EMBL Data Library, December 1997
A/Description: The sequence of C. elegans coemid F42A6.
A/Reference number: Z21201
A/Accession: T32619
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-182 <DUZ>
A/Cross-references: UNIPROT:Q44487; EMBL:AF038613; PIDN:AAB92052.1; GSPDB:GN00022; CESP:
A/Experimental source: strain Bristol N2; clone F42A6
C/Genetics:
A/Gene: CESP:F42A6.6
A/Map position: 4
A/Introns: 56/1; 106/2

Query Match          0.7%; Score 7; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 168 QAADGTQ 174
Db 164 QAADGTQ 170

RESULT 64
A64438
hypothetical protein MJ1106 - Methanococcus jannaschii
C/Species: Methanococcus jannaschii
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C/Accession: A64438
R/Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A/Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A/Reference number: A64300; MUID:96337999; PMID:8688087
A/Accession: A64438
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-185 <BUL>
A/Cross-references: UNIPROT:Q58506; GB:U67553; GB:L77117; NID:g1591744; PIDN:AAB99109.1;
C/Genetics:
A/Map position: REV1046608-1046051

Query Match          0.7%; Score 7; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 536 IYAIIVLP 542
Db 121 IYAIIVLP 127

RESULT 65
T09390
21K protein precursor - alfalfa
C/Species: Medicago sativa (alfalfa)
```

```
C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C/Accession: T09390
R/Coronado, C.
submitted to the EMBL Data Library, February 1997
A/Reference number: Z16658
A/Accession: T09390
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-187 <COR>
A/Cross-references: UNIPROT:P93674; EMBL:Y11553; NID:ei029850; PID:e304664
A/Experimental source: strain A2, subspecies varia; nodule
F:1-18/Domain: signal sequence #status Predicted <SIG>
F:19-187/Product: 21 K protein #status Predicted <MAT>

Query Match          0.7%; Score 7; DB 2; Length 187;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 SLSLNKT 127
Db 63 SLSLNKT 69

RESULT 66
I70176
dynein-like protein 6 - rat (fragment)
C/Species: Rattus norvegicus (Norway rat)
C/Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 09-Jul-2004
C/Accession: I70176
R/Tanaka, Y.; Zhang, Z.; Hirokawa, N.
J. Cell Sci. 108, 1883-1893, 1995
A/Title: Identification and molecular evolution of new dynein-like protein sequences in
A/Reference number: I55515; MUID:95386588; PMID:7657712
A/Accession: I70176
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-188 <RES>
A/Cross-references: UNIPROT:Q63169; GB:D26497; NID:g871908; PIDN:BAA05505.1; PID:g871909
C/Genetics:
A/Gene: DLP6

Query Match          0.7%; Score 7; DB 2; Length 188;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 276 LPDNLKA 282
Db 132 LPDNLKA 138

RESULT 67
E90187
hA1 protein [imported] - Sulfolobus solfataricus
C/Species: Sulfolobus solfataricus
C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C/Accession: E90187
R/She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Jensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A/Description: Sulfolobus solfataricus complete genome.
A/Reference number: A99139
A/Accession: E90187
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-192 <KUR>
A/Cross-references: UNIPROT:Q97ZZ0; GB:AE006641; NID:gi3813584; PIDN:AAK40756.1; GSPDB:G
C/Genetics:
A/Gene: SS00432
C/Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0226

Query Match          0.7%; Score 7; DB 2; Length 192;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 SDBLEI 504  
|||||

Db 44 SDBLEI 50  
|||||

RESULT 68  
E75462  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: E75462  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: E75462  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-195 <WHI>  
A:Cross-references: UNIPROT:Q9RV4; GB:AE001942; GB:AE000513; NID:G6458611; PIDN:AAF1047  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0897  
A:Map position: 1  
C:Superfamily: Deinococcus radiodurans hypothetical protein DR0897

Query Match 0.7%; Score 7; DB 2; Length 195;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 991 YTSVALL 997  
|||||

Db 116 YTSVALL 122  
|||||

RESULT 69  
F71246  
probable indolepyruvate ferredoxin oxidoreductase beta subunit - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 12-Jul-2004  
C:Accession: F71246  
R:Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Onfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: F71246  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-196 <KAW>  
A:Cross-references: UNIPROT:O57968; GB:AP000001; NID:G3236128; PIDN:BAA29301.1; PID:G325  
A:Experimental source: strain OT3  
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH0229

Query Match 0.7%; Score 7; DB 1; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 DLSEIKE 506  
|||||

Db 115 DLSEIKE 121  
|||||

RESULT 70  
T29084  
3-isopropylmalate dehydratase (EC 4.2.1.33) small chain - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor

C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T29084  
R:Redenbach, M.; Kieser, H.M.; Denapaita, D.; Eichner, A.; Cullum, J.; Kinaishi, H.; Hop  
Mol. Microbiol. 21, 77-96, 1996  
A:Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb  
A:Reference number: Z20556; MUID:97000351; PMID:8843436  
A:Accession: T29084  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-197 <RED>  
A:Cross-references: UNIPROT:O86535; EMBL:AL031124; PIDN:CAA20002.1  
C:Genetics:  
A:Note: leud  
C:Superfamily: 3-isopropylmalate dehydratase small chain  
C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 0.7%; Score 7; DB 2; Length 197;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 FRGSLK 327  
|||||

Db 101 FRGSLK 107  
|||||

RESULT 71  
T44358  
hypothetical protein ruva [imported] - Clostridium histolyticum  
C:Species: Clostridium histolyticum  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T44358  
R:Matsumura, O.; Jung, C.M.; Katayama, S.; Minami, J.; Takahashi, Y.; Okabe, A.  
J. Bacteriol. 181, 923-933, 1999  
A:Title: Gene duplication and multiplicity of collagenases in Clostridium histolyticum.  
A:Reference number: Z22752; MUID:99121032; PMID:9922257  
A:Accession: T44358  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-197 <MAT>  
A:Cross-references: UNIPROT:Q9ZJ6; EMBL:AB014075; NID:G3868863; PIDN:BAA34545.1; PID:G3  
A:Experimental source: strain JCM 1403  
C:Genetics:  
A:Note: ruva  
C:Superfamily: holliday junction DNA helicase ruva

Query Match 0.7%; Score 7; DB 2; Length 197;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 LELKDKL 523  
|||||

Db 126 LELKDKL 132  
|||||

RESULT 72  
A96504  
protein F9C16.19 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: A96504  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: A96504  
A>Status: preliminary

A;Molecule type: DNA  
A;Residues: 1-200 <STO>  
A;Cross-references: UNIPROT:Q9LNZ9; GB:AE005173; NID:g8778673; PIDN:AAF79681.1; GSPDB:GN  
C;Genetics:  
A;Gene: F9C16.19  
A;Map position: 1

Query Match 0.7%; Score 7; DB 2; Length 200;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 QASSSIS 43  
|||||  
Db 188 QASSSIS 194

RESULT 73  
I70171  
dynein-like protein 1 - rat (fragment)  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 09-Jul-2004  
A;Accession: I70171  
R;Tanaka, Y.; Zhang, Z.; Hirokawa, N.  
J. Cell Sci. 108, 1883-1893, 1995  
A;Title: Identification and molecular evolution of new dynein-like protein sequences in  
A;Reference number: 155515; MUID:95386588; PMID:7657712  
A;Accession: I70171  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-200 <RES>  
A;Cross-references: UNIPROT:Q63164; GB:D26492; NID:g871898; PIDN:BAA05500.1; PID:g871899  
C;Genetics:  
A;Gene: DUPI

Query Match 0.7%; Score 7; DB 2; Length 200;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 276 LPDNLKA 282  
|||||  
Db 144 LPDNLKA 150

RESULT 74  
G64006  
hypothetical protein HI0370 - Haemophilus influenzae (strain Rd KW20)  
C;Species: Haemophilus influenzae  
C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 09-Jul-2004  
A;Accession: G64006  
R;Pleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A;Reference number: A64000; MUID:95350630; PMID:7542800  
A;Accession: G64006  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-204 <TIGR>  
A;Cross-references: UNIPROT:P43989; GB:U32721; GB:L42023; NID:g1573334; PIDN:AAC22028.1;  
C;Superfamily: hypothetical protein HI0370

Query Match 0.7%; Score 7; DB 2; Length 204;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 SGSLEQ 117  
|||||  
Db 188 SGSLEQ 194

RESULT 75

S48952  
hypothetical protein YHR110w - yeast (Saccharomyces cerevisiae)  
C;Species: Saccharomyces cerevisiae  
C;Date: 13-Jan-1995 #sequence\_revision 10-Feb-1995 #text\_change 09-Jul-2004  
A;Accession: S48952  
R;Latreille, P.  
submitted to the EMBL Data Library, May 1994  
A;Description: The sequence of S. cerevisiae cosmid 8263.  
A;Reference number: S46676  
A;Accession: S48952  
A;Molecule type: DNA  
A;Residues: 1-212 <LAT>  
A;Cross-references: UNIPROT:P38819; EMBL:U00059; NID:g529116; PIDN:AAB68853.1; PID:g5291  
C;Genetics:  
A;Gene: SGD:ERP5  
A;Cross-references: SGD:S0001152  
A;Map position: 8R  
C;Superfamily: conserved hypothetical protein YHR110w  
C;Keywords: transmembrane protein

Query Match 0.7%; Score 7; DB 2; Length 212;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 950 SESANSK 956  
|||||  
Db 173 SESANSK 179

RESULT 76  
B84206  
hypothetical protein vng0475c [imported] - Halobacterium sp. NRC-1  
C;Species: Halobacterium sp. NRC-1  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
A;Accession: B84206  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.  
; Leithaeuser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabbo  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A;Title: Genome sequence of Halobacterium species NRC-1.  
A;Reference number: A84160; MUID:20504483; PMID:11016950  
A;Accession: B84206  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-214 <STO>  
A;Cross-references: UNIPROT:Q9HRZ4; GB:AE004437; NID:gi0580082; PIDN:AAG19014.1; GSPDB:G  
C;Genetics:  
A;Gene: VNG0475C

Query Match 0.7%; Score 7; DB 2; Length 214;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 883 AVATLAD 889  
|||||  
Db 167 AVATLAD 173

RESULT 77  
G95069  
ABC transporter, ATP-binding protein Vexp2 SP0600 [imported] - Streptococcus pneumoniae  
C;Species: Streptococcus pneumoniae  
C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
A;Accession: G95069  
R;Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
naon, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A;Reference number: A95000; MUID:21357209; PMID:11463916  
A;Accession: G95069

A;Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-215 <KUR>  
A;Cross-references: UNIPROT:Q9S4J1; GB:AE005672; PIDN:AAK74752.1; PID:g14972074; GSPDB:G  
A;Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0600

Query Match 0.7%; Score 7; DB 2; Length 215;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKDK 522  
DB 199 TLELKDK 205  
|||||

RESULT 78  
E97937  
hypothetical protein vex2 [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: E97937  
R:Hosking, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: E97937  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-215 <KUR>  
A;Cross-references: UNIPROT:Q8DQ83; GB:AE007317; PIDN:AAK99329.1; PID:g15458099; GSPDB:G  
A:Gene: vex2

Query Match 0.7%; Score 7; DB 2; Length 215;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKDK 522  
DB 199 TLELKDK 205  
|||||

RESULT 79  
C75637  
hypothetical protein DRC0023 - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: C75637  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: C75637  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-219 <WHI>  
A;Cross-references: UNIPROT:Q9RZG4; GB:AE001827; NID:g6460959; PIDN:AAF12691.1; PID:g646  
A;Experimental source: strain R1  
C:Genetics:  
A:Gene: DRC0023  
A:Map position: plasmid  
A:Genome: plasmid  
A:Note: plasmid CPl  
C:Superfamily: Deinococcus radiodurans hypothetical protein DRC0023

Query Match 0.7%; Score 7; DB 2; Length 219;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LEEVKLP 476  
DB 84 LEEVKLP 90  
|||||

RESULT 80  
A64479  
DNA-(apurinic or apyrimidinic site) lyase (EC 4.2.99.18) - Methanococcus jannaschii  
N:Alternate names: endonuclease III  
C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
C:Accession: A64479  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
rson, J.D.; Sadov, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A:Reference number: A64300; MUID:96337999; PMID:8688087  
A:Accession: A64479  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-220 <BUL>  
A;Cross-references: UNIPROT:Q58829; GB:U67584; GB:L77117; NID:g1592077; PIDN:AA899444.1;  
C:Genetics:  
A:Map position: FOR1403656-1404318  
C:Superfamily: Methanococcus jannaschii conserved DNA-(apurinic or apyrimidinic site) lyase  
C:Keywords: carbon-oxygen lyase

Query Match 0.7%; Score 7; DB 2; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 869 ILNVED 875  
DB 67 ILNVED 73  
|||||

RESULT 81  
JQ1986  
avrPmaA1 protein - Pseudomonas syringae  
C:Species: Pseudomonas syringae  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C:Accession: JQ1986; S33383  
R:Dangl, J.L.; Ritter, C.; Gibbon, M.J.; Mur, L.A.J.; Wood, J.R.; Goss, S.; Mansfield, J  
Plant Cell 4, 1359-1369, 1992  
A:Title: Functional homologs of the Arabidopsis RPM1 disease resistance gene in bean and  
A:Reference number: JQ1985; MUID:93120881; PMID:1477552  
A:Accession: JQ1986  
A:Molecule type: DNA  
A:Residues: 1-220 <DAN>  
A;Cross-references: UNIPROT:Q52537; GB:X67808; NID:g45863; PIDN:CAA48009.1; PID:g871512  
A;Experimental source: strain P.S.pv maculicola  
C:Genetics:  
A:Gene: avrPmaA1

Query Match 0.7%; Score 7; DB 2; Length 220;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ENGEISQ 206  
DB 154 ENGEISQ 160  
|||||

RESULT 82  
JQ1985  
avrPpaI protein - Pseudomonas syringae  
C:Species: Pseudomonas syringae  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004

C/Accession: JQ1985; S33384  
 R;Dargl, J.L.; Ritter, C.; Gibbon, M.J.; Mur, L.A.J.; Wood, J.R.; Goss, S.; Mansfield, J.  
 Plant Cell 4, 1359-1369, 1992  
 A/Title: Functional homologs of the Arabidopsis RPM1 disease resistance gene in bean and  
 A/Reference number: JQ1985; MUID:93120881; PMID:1477552  
 A/Accession: JQ1985  
 A/Molecule type: DNA  
 A/Residues: 1-220 <DNA>  
 A/Cross-references: UNIPROT:Q52545; GB:X67807; NID:G45864; PIDN:CAA48008.1; PID:G296311  
 A/Experimental source: strain P.S.pv pisi race2  
 C/Genetics:  
 A/Gene: avrPp1A1

Query Match 0.7%; Score 7; DB 2; Length 220;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ENGEISQ 206  
 Db 154 ENGEISQ 160  
 |||||

RESULT 83  
 H83639  
 hypothetical protein PA0045 [imported] - Pseudomonas aeruginosa (strain PAO1)  
 C/Species: Pseudomonas aeruginosa  
 C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
 C/Accession: H83639  
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho  
 A/Reference number: A82950; MUID:20437337; PMID:10984043  
 A/Accession: H83639  
 A/Molecule type: DNA  
 A/Status: preliminary  
 A/Residues: 1-228 <STO>  
 A/Cross-references: UNIPROT:Q91787; GB:AE004444; GB:AE004091; NID:G9945863; PIDN:AAG0343  
 A/Experimental source: strain PAO1  
 C/Genetics:  
 A/Gene: PA0045

Query Match 0.7%; Score 7; DB 2; Length 228;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 882 LAVATILA 888  
 Db 13 LAVATILA 19  
 |||||

RESULT 84  
 C84606  
 probable RAS type GTP-binding protein [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
 C/Accession: C84606  
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A/Reference number: A84420; MUID:20083487; PMID:10617197  
 A/Accession: C84606  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-230 <STO>  
 A/Cross-references: UNIPROT:Q9SJ11; GB:AE002093; NID:G4417298; PIDN:AAD20423.1; GSPDB:GN  
 C/Genetics:  
 A/Gene: At2g21880  
 A/Map position: 2  
 C/Superfamily: ras transforming protein; translation elongation factor Tu homology

Query Match 0.7%; Score 7; DB 2; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 490 NNLSKPE 496  
 Db 92 NNLSKPE 98  
 |||||

RESULT 85  
 JQ1601  
 replication protein C - Clostridium butyricum plasmid pCB101  
 C/Species: Clostridium butyricum  
 C/Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Jul-2004  
 C/Accession: JQ1601; S18114  
 R;Brehm, J.K.; Pennock, A.; Bullman, H.M.S.; Young, M.; Oultram, J.D.; Minton, N.P.  
 Plasmid 28, 1-13, 1992  
 A/Title: Physical characterization of the replication origin of the cryptic plasmid pCB1  
 A/Reference number: PQ0372; MUID:92390513; PMID:1518909  
 A/Accession: JQ1601  
 A/Molecule type: DNA  
 A/Residues: 1-231 <BRE>  
 A/Cross-references: UNIPROT:Q45961; EMBL:X62684; NID:G40399; PIDN:CAA44561.1; PID:G40402  
 A/Experimental source: strain NCIB7423  
 C/Comment: This protein is required for autonomous replication of the plasmid.  
 C/Genetics:  
 A/Gene: repC  
 A/Genome: plasmid  
 C/Keywords: plasmid replication

Query Match 0.7%; Score 7; DB 2; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 902 KLSQTS 908  
 Db 11 KLSQTS 17  
 |||||

RESULT 86  
 C72086  
 sam-dependent methyltransferase - Chlamydia pneumoniae (strain CWL029)  
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
 C/Accession: C72086  
 R;Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
 Nature Genet. 21, 385-389, 1999  
 A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A/Reference number: A72000; MUID:99206606; PMID:10192388  
 A/Accession: C72086  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-238 <ARN>  
 A/Cross-references: UNIPROT:Q9Z8G1; GB:AE001622; GB:AE001363; NID:G4376652; PIDN:AAD1852  
 A/Experimental source: strain CWL029  
 C/Genetics:  
 A/Gene: yabc 1  
 C/Superfamily: methyltransferase, YraI type

Query Match 0.7%; Score 7; DB 2; Length 238;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 595 IPVQAFS 601  
 Db 111 IPVQAFS 117  
 |||||

RESULT 87  
 DB6538  
 SAM-dependent methyltransferase [imported] - Chlamydia pneumoniae (strain J138)  
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
 C:Accession: D86538  
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ie  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A:Reference number: A86491; MUID:20330349; PMID:10871362  
 A:Accession: D86538  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-238 <STO>  
 A:Cross-references: UNIPROT:Q9TRJ3; GB:BA000008; NID:g8978754; PIDN:BAA98590.1; GSPDB:GN  
 A:Experimental source: strain J138  
 C:Genetics:  
 C:Superfamily: methyltransferase, YxaL type

Query Match 0.7%; Score 7; DB 2; Length 238;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 595 IPVQAPS 601  
 |||||  
 DB 111 IPVQAPS 117

RESULT 88  
 H64343  
 hypothetical protein M30352 - Methanococcus jannaschii  
 C:Species: Methanococcus jannaschii  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
 R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
 ; Reich, C.I.; Overbeck, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;  
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273, 1038-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C  
 A:Title: Complete genome sequence of the methanogenic archaeson, Methanococcus jannaschii  
 A:Reference number: A64300; MUID:96337999; PMID:8688087  
 A:Accession: H64343  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-239 <BUL>  
 A:Cross-references: UNIPROT:Q57798; GB:U67488; GB:L77117; NID:g2826273; PIDN:AAB98341.1;  
 C:Genetics:  
 A:Map position: REV3233901-323182

Query Match 0.7%; Score 7; DB 2; Length 239;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 ETLELKD 521  
 |||||  
 DB 50 ETLELKD 56

RESULT 89  
 A86719  
 oxidoreductase yhgA [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
 C:Species: Lactococcus lactis subsp. lactis  
 C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004  
 C:Accession: A86719  
 R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissensbach, J.; Ehrli  
 Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
 A:Reference number: A86625; MUID:21235186; PMID:11337471  
 A:Accession: A86719  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-242 <STO>  
 A:Cross-references: UNIPROT:Q9CHH8; GB:AE005176; PID:g12723669; PIDN:AAK04851.1; GSPDB:G  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 C:Superfamily: yhgA

C:Superfamily: NADPH-flavin oxidoreductase homolog

Query Match 0.7%; Score 7; DB 2; Length 242;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 687 DLEKILK 693

DB 29 DLEKILK 35

RESULT 90

A85688  
 hypothetical protein 21868 [imported] - Escherichia coli (strain O157:H7, substrain EDL)  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
 C:Accession: A85688  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: A85688  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-247 <STO>  
 A:Cross-references: UNIPROT:O8X3P1; GB:AE005174; NID:g12514793; PIDN:AAG55965.1; GSPDB:  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 C:Superfamily: phage P22 gene 18 protein

Query Match 0.7%; Score 7; DB 2; Length 247;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 EVNLPEP 257

DB 96 EVNLPEP 102

RESULT 91

A84653  
 hypothetical protein At2g25810 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 22-Mar-2002  
 C:Accession: A84653  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: A84653  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-249 <STO>  
 A:Cross-references: GB:AE002093; NID:g3643602; PIDN:AAC42249.1; GSPDB:GN00139  
 C:Genetics:  
 A:Map position: 2  
 C:Superfamily: nodulin-26

Query Match 0.7%; Score 7; DB 2; Length 249;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 GNTLVGL 149

DB 47 GNTLVGL 53

RESULT 92



```

AB2991
transcription regulator, GntR family Atu3532 [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AB2991
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Gillet, M.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2991
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <KUR>
A:CROSS-references: UNIPROT:Q8UA43; GB:AE008689; PIDN:AA44344.1; PID:gi7741936; GSPDB:C
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3532
A:Map position: linear chromosome

Query Match          0.7%; Score 7; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 755 AEKALVT 761
|||||
Db 221 AEKALVT 227

RESULT 93
G98292
hypochemical transcription regulator in dcp-nohA intergenic region [imported] - Agrobact
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: G98292
R:Goodner, B.; Hipkile, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: G98292
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <KUR>
A:CROSS-references: UNIPROT:Q8UA43; GB:AE007870; PIDN:AAK89865.1; PID:gi5159808; GSPDB:C
C:Genetics:
A:Gene: AGR_L_2599
A:Map position: linear chromosome

Query Match          0.7%; Score 7; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 755 AEKALVT 761
|||||
Db 221 AEKALVT 227

RESULT 94
AB6727
transcription regulator yida [imported] - Lactococcus lactis subsp. lactis (strain IL140
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: AB6727
R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis se
A:Reference number: AB6625; MUID:21235186; PMID:11337471
A:Accession: AB6727
A>Status: preliminary

```

```

A:Molecule type: DNA
A:Residues: 1-251 <STO>
A:CROSS-references: UNIPROT:Q9CHB5; GB:AE005176; PID:gl2723740; PIDN:AAK04915.1; GSPDB:G
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yida

Query Match          0.7%; Score 7; DB 2; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 591 QLTEIPV 597
|||||
Db 73 QLTEIPV 79

RESULT 95
A96761
unknown protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A96761
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96761
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <STO>
A:CROSS-references: UNIPROT:Q9FDZ8; GB:AE005173; NID:gl1120794; PIDN:AAK30974.1; GSPDB:G
C:Genetics:
A:Gene: T9L24.51
A:Map position: 1

Query Match          0.7%; Score 7; DB 2; Length 254;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 747 DLDKXIA 753
|||||
Db 144 DLDKXIA 150

RESULT 96
WNAD84
early E1A 28K protein - human adenovirus 4
C:Species: Mastadenovirus h4 (human adenovirus 4)
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C:Accession: A25614
R:Tokunaga, O.; Yaegashi, T.; Lowe, J.; Dobbs, L.; Padmanabhan, R.
Virology 155, 418-433, 1986
A:Title: Sequence analysis in the E1 region of adenovirus type 4 DNA.
A:Reference number: A94347; MUID:87071662; PMID:2947381
A:Accession: A25614
A:Molecule type: DNA
A:Residues: 1-257 <TKO>
A:CROSS-references: UNIPROT:P10407; GB:M14918; NID:g209874; PIDN:AAA67091.1; PID:g825435
C:Genetics:
A:Insertions: 189/2
C:Superfamily: adenovirus early E1A protein
C:Keywords: DNA binding; early protein; transcription regulation; zinc finger

Query Match          0.7%; Score 7; DB 1; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 924 QAIQNAA 930  
|||||

Db 129 QAIQNAA 135

RESULT 97

AH1865  
hypothetical protein all0473 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AH1865  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH1865  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-257 <KUR>  
A:Cross-references: UNIPROT:Q8Y217; GB:BA000019; PIDN:BAE72431.1; PID:gl7129818; GSPDB:G0000019  
C:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all0473  
C:Superfamily: gufa protein

Query Match 0.7%; Score 7; DB 2; Length 257;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 801 LLTGLVE 807  
|||||

Db 182 LLTGLVE 188

RESULT 98

S36431  
hypothetical protein - Buchnera aphidicola  
C:Species: Buchnera aphidicola  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
C:Accession: S36431  
R: Munson, M.A.; Baumann, P.  
submitted to the EMBL Data Library, December 1992  
A:Description: tryptophan operon of Buchnera aphidicola.  
A:Reference number: S36426  
A:Accession: S36431  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-258 <MUN>  
A:Cross-references: UNIPROT:P42395; EMBL:Z19055; NID:g396655; PIDN:CAA79503.1; PID:g396655  
C:Superfamily: conserved hypothetical protein H1243

Query Match 0.7%; Score 7; DB 2; Length 258;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 430 VKRKNKL 436  
|||||

Db 59 VKRKNKL 65

RESULT 99

T36359  
hypothetical protein SCE9.40 - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T36359  
R: Oliver, K.; Harrie, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: Z21575

A:Accession: T36359  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-258 <OLI>  
A:Cross-references: UNIPROT:Q9X8L2; EMBL:AL049841; PIDN:CAB42786.1; GSPDB:GN00070; SCORE:100.0  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOE9.SCE9.40

Query Match 0.7%; Score 7; DB 2; Length 258;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 999 LITAICK 1005  
|||||

Db 16 LITAICK 22

RESULT 100

T19091  
hypothetical protein C08F11.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T19091  
R: Matthews, L.  
submitted to the EMBL Data Library, December 1996  
A:Reference number: Z19072  
A:Accession: T19091  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-259 <WIL>  
A:Cross-references: UNIPROT:O62050; EMBL:Z83216; PIDN:CAB05675.1; GSPDB:GN00022; CESP:C08F11  
A:Experimental source: clone C08F11  
C:Genetics:  
A:Gene: CESP:C08F11.4  
A:Map position: 4  
A:Introns: 81/3; 132/3; 202/3  
C:Superfamily: Caenorhabditis elegans hypothetical protein C33A12.9b

Query Match 0.7%; Score 7; DB 2; Length 259;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LTLTTVS 16  
|||||

Db 184 LTLTTVS 190

Search completed: August 28, 2005, 11:08:51  
Job time : 53 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:29:30 ; Search time 173 Seconds  
(without alignments)  
2253.494 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 5080

Sequence: 1 MKKHLKTVALLTTSVVTH.....LGYTSVALLSLTAIKKKY 1008

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 385760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 110 summaries

Database : A\_Geneseq\_16Dec04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	5080	100.0	1008	5	Abp25813 Streptococ
2	5080	100.0	1008	5	Abp53347 Streptococ
3	5080	100.0	1008	8	Adr83923 S. pyogen
4	3777.5	74.4	1055	4	Aau03612 Group B S
5	3777.5	74.4	1055	6	Abp56257 Serotype
6	3776.5	74.3	1055	5	Abp25812 Streptococ
7	243	4.8	4734	8	Adh62807 Lactobaci
8	241.5	4.8	669	8	Adh22318 Bacteri
9	237.5	4.7	668	8	Adh22319 Bacteri
10	220.5	4.3	3257	4	Abp67502 Drosophil
11	217.5	4.3	1072	5	Abp54963 Lactococc
12	215.5	4.2	641	8	Adk47748 Streptococ
13	215.5	4.2	641	8	Adr94313 Novel S.
14	214.5	4.2	3225	7	Adj68448 Human hea
15	214.5	4.2	3259	7	Adh56037 Human Pro
16	214.5	4.2	3259	7	Adh56033 Human Pro
17	213.5	4.2	621	6	Abu02658 S. pneumo
18	213.5	4.2	621	6	Abp81493 Streptococ
19	213.5	4.2	621	8	Adh92258 S. pneumon
20	213	4.2	1799	7	Adh87293 Enterococ
21	213	4.2	2045	6	Abu46539 Protein e
22	213	4.2	2045	6	Adh83918 S. pyogen
23	213	4.2	2059	5	Abp25711 Streptococ
24	211	4.2	1881	5	Abp73809 Candida a
25	209.5	4.1	1254	2	Aar07503 Merozoite

Aaw24575 Merozoite  
Adi45155 Rice isop  
Abu42253 Protein e  
Abm73008 Staphyloc  
Abj18935 Pathogen  
Aau34338 Staphyloc  
Aau37403 Staphyloc  
Abj19119 Pathogen  
Aag82935 S. epider  
Abu42656 Protein e  
Abp38314 Staphyloc  
Adh4851 S epiderm  
Adh04803 Staphyloc  
Abu15838 Protein e  
Abi62804 Methicill  
Abu15887 Protein e  
Aau34207 Staphyloc  
Aau37018 Staphyloc  
Aay19935 B. burgdo  
Aay19934 B. burgdo  
Aar46605 Malarial  
Aaw98149 Bacillus  
Adh83955 S. pyogen  
Abp62805 Methicill  
Abp56875 Staphyloc  
Abi62792 MRSA FmtB  
Abb63519 Drosophil  
Abi53116 Protein s  
Adh63040 Disease t  
Adh18933 Bacteri  
Aau34139 Staphyloc  
Aau36796 Staphyloc  
Aab18161 Plasmodiu  
Aar46608 Plasmodiu  
Abp53560 Protein s  
Adh64380 Disease t  
Adh43855 Bacteri  
Abb47314 Listeria  
Abu33037 Protein e  
Abp40235 Staphyloc  
Adh06647 Staphyloc  
Aau35908 Helicobac  
Adh62806 Lactobaci  
Aau37320 Staphyloc  
Abm71190 Staphyloc  
Adh62805 Lactobaci  
Aau34339 Staphyloc  
Abj19059 Pathogen  
Aar72826 Human mit  
Aau23996 Human mit  
Adh95402 Human BSC  
Adk70220 Human oss  
Adi72172 Human sol  
Adi18045 Human sof  
Aau99795 Kinetocho  
Adh084842 E faecal  
Abb60349 Drosophil  
Adf45530 Chicken A  
Abu43311 Protein e  
Adh22762 Bacteri  
Adh19417 Bacteri  
Abu42252 Protein e  
Abp29526 Streptoco  
Abb47751 Listeria  
Abu32684 Protein e  
Adh89551 Staphyloc  
Aau37374 Staphyloc  
Aau34320 Staphyloc  
Abj19002 Pathogen  
Abm71899 Staphyloc  
Abu07438 Protein d  
Adj66483 Cenp-F ki  
Abb63571 Drosophil

99 186.5 3.7 1312 5 ABB77984  
 100 186.5 3.7 1312 6 ABR53887  
 101 186.5 3.7 1312 7 ADK63602  
 102 186.5 3.7 1312 8 ADN19297  
 103 185.5 3.7 1186 8 ADS44694  
 104 185.5 3.7 1788 6 ABU18740  
 105 185 3.6 2086 4 AAU34143  
 106 185 3.6 3818 8 ADP45525  
 107 185 3.6 3899 6 ABR92048  
 108 185 3.6 3900 8 ADP45500  
 109 185 3.6 3907 6 ABR92047  
 110 185 3.6 3908 8 ADP45502

## ALIGNMENTS

RESULT 1  
 ABP25813  
 ID ABP25813 standard; protein; 1008 AA.  
 XX  
 AC ABP25813;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polypeptide SEQ ID NO 802.  
 XX  
 KW Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX  
 OS Streptococcus pyogenes.  
 XX  
 PN WO200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-GB004789.  
 XX  
 PR 27-OCT-2000; 2000GB-00026333.  
 PR 24-NOV-2000; 2000GB-00028727.  
 PR 07-MAR-2001; 2001GB-00005640.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 DR WPI, 2002-352536/38.  
 DR N-PSDB; ABN66444.  
 XX

New Streptococcus protein for the treatment or prevention of infection or disease caused by Streptococcus bacteria, such as meningitis, and for detecting a compound that binds to the protein.

Claim 1; Page 3231; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in the specification. The proteins have antibacterial and antiinflammatory activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly S. agalactiae and S. pyogenes. Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins  
 XX  
 SQ Sequence 1008 AA;  
 Query Match 100.0%; Score 5080; DB 5; Length 1008;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-301; Indels 0; Gaps 0;  
 Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKKHLKVALTLTVSVVTHNQEVFSLVKSPILKQTQASSISGADYAESGSKLKINE 60  
 DB 1 MKKHLKVALTLTVSVVTHNQEVFSLVKSPILKQTQASSISGADYAESGSKLKINE 60  
 QY 61 TSGPVDVTVDLPSDKRTTPEKIKONLAKGPREOELKAVTENTSEKQITSGSOLEQSK 120  
 DB 61 TSGPVDVTVDLPSDKRTTPEKIKONLAKGPREOELKAVTENTSEKQITSGSOLEQSK 120  
 QY 121 SLSLNTKVPSTSNWEICDFTKGNLTVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVAS 180  
 DB 121 SLSLNTKVPSTSNWEICDFTKGNLTVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVAS 180  
 QY 181 FAFTPDKKTAAIAYTSRAGENGESQLDVGKEIINEGEVFNLSLLKKVTIPTGYKHIGQ 240  
 DB 181 FAFTPDKKTAAIAYTSRAGENGESQLDVGKEIINEGEVFNLSLLKKVTIPTGYKHIGQ 240  
 QY 241 DAFVNDKNIAEAVNLPSLETISDYAPAHALAKQIDLPDNLKAIGELAFDQITGKLSLP 300  
 DB 241 DAFVNDKNIAEAVNLPSLETISDYAPAHALAKQIDLPDNLKAIGELAFDQITGKLSLP 300  
 QY 301 RQLMLRAERAPKSNHIIKTIEPRGNSLVKIGASQDNDLSQLMLPDGLEKIESAFTGNP 360  
 DB 301 RQLMLRAERAPKSNHIIKTIEPRGNSLVKIGASQDNDLSQLMLPDGLEKIESAFTGNP 360  
 QY 361 GDDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420  
 DB 361 GDDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420  
 QY 421 GFSNKGLOKVKRNQNLKLEIPKOHNGVTITEIGDNAFRNVDQNKTLRKYDLEEVKLPSTIR 480  
 DB 421 GFSNKGLOKVKRNQNLKLEIPKOHNGVTITEIGDNAFRNVDQNKTLRKYDLEEVKLPSTIR 480  
 QY 481 KIGAPAFQSNLKSFEASDDLEEKEGAFNNRITETLELKDQKVTIGDAAAPHIHIYAIV 540  
 DB 481 KIGAPAFQSNLKSFEASDDLEEKEGAFNNRITETLELKDQKVTIGDAAAPHIHIYAIV 540  
 QY 541 LPESVQEIGRSAPRQNGANNLIIPGSKVKTILGEMAFLSNRLHLDLSEKQLTPEIQAQF 600  
 DB 541 LPESVQEIGRSAPRQNGANNLIIPGSKVKTILGEMAFLSNRLHLDLSEKQLTPEIQAQF 600  
 QY 601 SDNALKEVLLPASLTKTIREAFKQHLKQLEVASALSIAFNALDDNDGDQFQNKVVVK 660  
 DB 601 SDNALKEVLLPASLTKTIREAFKQHLKQLEVASALSIAFNALDDNDGDQFQNKVVVK 660  
 QY 661 THNSYALADGEHFIVDPDKLSSTIVLEKILKLEGLDYSTLTQTTQTFQFRDMMTAGKA 720  
 DB 661 THNSYALADGEHFIVDPDKLSSTIVLEKILKLEGLDYSTLTQTTQTFQFRDMMTAGKA 720  
 QY 721 LLSKSNLRQGEKQKFLQEAQPFGLGRVLDKAIAKAELVTKKATQNGQLLERSINKAVL 780  
 DB 721 LLSKSNLRQGEKQKFLQEAQPFGLGRVLDKAIAKAELVTKKATQNGQLLERSINKAVL 780  
 QY 781 AYNNSAIKCANVKLEKELDLTLTGLVEGKGPLAQATWQGVYLLKPLPEYVIGLNVY 840  
 DB 781 AYNNSAIKCANVKLEKELDLTLTGLVEGKGPLAQATWQGVYLLKPLPEYVIGLNVY 840  
 QY 841 FDKSGKLIYALDMSDTTIGEGQDAYGNPILNVDNEDNHYHALAVATLADYEGLDIKITLN 900  
 DB 841 FDKSGKLIYALDMSDTTIGEGQDAYGNPILNVDNEDNHYHALAVATLADYEGLDIKITLN 900  
 QY 901 SKLSQLTSSIRQVPTAAVHRAGIPQAIQNAAEAEQLLPKPTGTHSEKSSSSSANSKDRGL 960  
 DB 901 SKLSQLTSSIRQVPTAAVHRAGIPQAIQNAAEAEQLLPKPTGTHSEKSSSSSANSKDRGL 960

QY 961 QSNPKTNRGRHSAILPRGSGSFVYGLGYTSVALLSLITAIKXKY 1008  
 DB |||||||  
 XX 961 QSNPKTNRGRHSAILPRGSGSFVYGLGYTSVALLSLITAIKXKY 1008

## RESULT 2

ABP53347

ID ABP53347 standard; protein; 1008 AA.

XX

AC ABP53347;

XX

DT 18-NOV-2002 (first entry)

XX

DE Streptococcus pyogenes BVH-P7 protein SEQ ID NO:2.

XX

KW Streptococcus pyogenes; BVH-P7; antigen; group A Streptococcus; GAS;  
 bacterial pathogen; vaccine; antibacterial; gene therapy; pharyngitis;  
 streptococcal infection; erysipelas; impetigo; scarlet fever; infection;  
 invasive disease; bacteraemia; necrotising fasciitis; toxic shock.

XX

OS Streptococcus pyogenes.

XX

FH Key Location/Qualifiers

FT Peptide 1..21

FT /label= signal

FT Protein 22..1008

FT /label= mature\_BVH-P7

PN

XX WO200266650-A2.

XX

PD 29-AUG-2002.

XX

PF 21-FEB-2002; 2002WO-CA000207.

XX

PR 21-FEB-2001; 2001US-0269840P.

XX

PA (SHIR-) SHIRE BIOCHEM INC.

XX

PI Martin D, Rioux S, Brodeur BR, Hamel J, Rheault P;

XX

PI WPI, 2002-674948/72.

XX

DR N-FSDB; ASQ81821.

XX

PT New polypeptide useful as a vaccine component for preventing, treating or  
 diagnosing Streptococcus pyogenes infections, e.g. pharyngitis.

XX

PT erysipelas, impetigo, scarlet fever, bacteremia, necrotizing fasciitis or  
 toxic shock.

XX

PS Claim 17; Fig 2; 52pp; English.

XX

CC The present sequence represents Streptococcus pyogenes BVH-P7 protein  
 (1). (1) has antibacterial activity and can be used in vaccines and gene  
 therapy. The Streptococcus pyogenes BVH-P7 polypeptide is useful as a  
 vaccine component for preventing, treating and/or diagnosing  
 streptococcal infections, such as pharyngitis, erysipelas, impetigo,  
 scarlet fever, and invasive diseases such as bacteraemia and necrotising  
 fasciitis, or toxic shock. A composition comprising the BVH-P7  
 polypeptide is useful in the manufacture of a medicament for the  
 prophylactic or therapeutic treatment of streptococcal infection. The BVH  
 -P7 polynucleotide may be used in designing probes for the detection of  
 Streptococcus in biological samples. The BVH-P7 polypeptide may also be  
 used as immunogens for the production of antibodies against streptococcal  
 infections

XX

XX Sequence 1008 AA;

Query Match 100.0%; Score 5080; DB 5; Length 1008;

Best Local Similarity 100.0%; Pred. No. 5.6e-301;

Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGHLKTVALLTLTVSVVTHNQVPSLVKPEILKOTQSSISGADYAESGSKLKINE 60

DB 1 MKGHLKTVALLTLTVSVVTHNQVPSLVKPEILKOTQSSISGADYAESGSKLKINE 60

QY 61 TSGPVDDTVTDLFSDKRTTPEKIKONLAKGPREQELKAVTENTESEKQITSGSOLEQSK 120  
 DB |||||||  
 QY 121 SLSLNKTVPSTSNWBEICDFITKGNLTGLSGSGVEKLSQTDHLVLPQAAADGTQIQVAS 180  
 DB |||||||  
 QY 181 FAFTPDKKTATAEYTSRAGENGESQLDVGDKGIIINEGEVFNYSLLKKVTITPTGFKHIGQ 240  
 DB |||||||  
 QY 241 DAFVNDKNIAEVLNPESLETISDYAFALHAKQLDLPDLNKAIGELAFDQITGKLSLP 300  
 DB |||||||  
 QY 301 RQLMRLAERAFKSNHIKTIIFRGNLSKLVIGASFDQNDLSQLMLPDGLEKIESEAFPTGNP 360  
 DB |||||||  
 QY 361 GDDHYNNRVLVLTGSKGNPSGLATENTYVNDPKSLWQESPEIDYTKWLEEDFTYKNSVT 420  
 DB |||||||  
 QY 421 GFSNKGKLVKRNKNLEIPKQHGVTITEIGDNAFRNVDFQNKTLRKYDLLEVKLPSTIR 480  
 DB |||||||  
 QY 481 KIGAPAFOSNNLKSFEASDDLEETKEGAFNMNRITETLEKOKLVITGDAAPHINHIYIV 540  
 DB |||||||  
 QY 541 LPESVQETIGRSAFRONGANNLIEMGSKVKTIGEMAFNSNRLEHLDLSEKOLTEIPVOAF 600  
 DB |||||||  
 QY 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSIAFNALDDNDGDEQFNKVVVK 660  
 DB |||||||  
 QY 661 THNSYALADGEHFTVDPDKLSSTIVDEKILKIEGLDYSTLRTOTQTFRDMTTAGKA 720  
 DB |||||||  
 QY 721 LLSKSNLRQGEKQFLQEAQFPLGRVDLDKAIAKAEKALVTCKATNGOLLERSINKAVL 780  
 DB |||||||  
 QY 781 AYNSAIKKNVKKRLEKELDLITGLVEGKPLAQATWVGVLKTLPLPEYIIGLVNY 840  
 DB |||||||  
 QY 841 FDKSGKLIYALDMSDTTIGEGQDAYGNPILNVNDEGECYHALAVATLADYGLDKITILN 900  
 DB |||||||  
 QY 901 SKLSQLSIRQVPTAAHYRAGIFQAIQWAAAEAEQLLPKPGTHSEKSSSSSANSKDRGL 960  
 DB |||||||  
 QY 961 QSNPKTNRGRHSAILPRGSGSFVYGLGYTSVALLSLITAIKXKY 1008  
 DB |||||||

## RESULT 3

ADR83923

ID ADR83923 standard; protein; 1008 AA.

XX

AC ADR83923;

XX

DT 02-DEC-2004 (first entry)



XX S. pyogenes hyperimmune system reactive antigen Spy0843.  
 XX hyperimmune serum reactive antigen; vaccine; anticaline.  
 XX Streptococcus pyogenes.  
 XX WO2004078907-A2.  
 XX 16-SEP-2004.  
 XX 02-MAR-2004; 2004WO-EP002087.  
 XX 04-MAR-2003; 2003EP-00450061.  
 XX (INTE-) INTERCELL AG.  
 XX Meinke A, Nagy E, Winkler B, Gelbmann D;  
 XX MPI; 2004-653698/63.  
 XX N-PSDB; ADR83773.  
 XX New isolated nucleic acid molecules encoding hyperimmune serum-reactive  
 XX antigens from Streptococcus pyogenes, useful for diagnosing, preventing  
 XX and treating S. pyogenes infections.  
 XX Claim 14; SEQ ID NO 191; 145pp; English.  
 XX This invention describes a novel nucleic acid molecule encoding a  
 XX hyperimmune serum reactive antigen or its fragment from Streptococcus  
 XX pyogenes. The nucleic acid molecule or hyperimmune serum-reactive antigen  
 XX or its fragment are useful for the manufacture of a pharmaceutical  
 XX preparation, especially a vaccine, against S. pyogenes infection. In  
 XX addition, the hyperimmune serum reactive antigen or fragment is used for  
 XX the isolation and/or purification and/or identification of an interaction  
 XX partner of the hyperimmune serum reactive antigen or its fragment, for  
 XX the generation of a peptide (e.g. anticalines) binding to the antigen or  
 XX fragment, or for the manufacture of a functional nucleic acid selected  
 XX from aptamers and spiegelmers. The nucleic acid molecule may also be used  
 XX for the manufacture of functional ribonucleic acids, such as ribozymes,  
 XX antisense nucleic acids and siRNA. ADR83733-ADR84189 represent S.  
 XX pyogenes hyperimmune serum reactive antigens, fragments and the encoding  
 XX polynucleotide described in the invention.  
 XX Sequence 1008 AA;

Query Match 100.0%; Score 5080; DB 8; Length 1008;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-301;  
 Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGHLKTVALLTLTVSVVTHNQEVPSLVKEPILKQTQASSISGADYAESGSKLKINE 60  
 DB 1 MKGHLKTVALLTLTVSVVTHNQEVPSLVKEPILKQTQASSISGADYAESGSKLKINE 60

QY 61 TSGPVDVDTDLPSDKRTTPKIKDNLAKPREQELKAVTENSEKQITSGSOLEOSKE 120  
 DB 61 TSGPVDVDTDLPSDKRTTPKIKDNLAKPREQELKAVTENSEKQITSGSOLEOSKE 120

QY 121 SLSLNTKVPSTSNWEICDFTKGNLTGLVLSGSGVEKLSQTDHLVLPQAADGTQLIOVAS 180  
 DB 121 SLSLNTKVPSTSNWEICDFTKGNLTGLVLSGSGVEKLSQTDHLVLPQAADGTQLIOVAS 180

QY 181 FAFTPDKKTATAEYTSRAGENGEISQLDVGKGIINEGEVFNYSLLKKVTIPTGYKHIGQ 240  
 DB 181 FAFTPDKKTATAEYTSRAGENGEISQLDVGKGIINEGEVFNYSLLKKVTIPTGYKHIGQ 240

QY 241 DAFVNDKNIAENVLPSELETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSLP 300  
 DB 241 DAFVNDKNIAENVLPSELETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSLP 300

QY 301 RQLMLAERAFKSNHIKTIIEFRGSLKVIIGASFDNDLSQLMLPDGLEKIESAFTGNP 360  
 DB 301 RQLMLAERAFKSNHIKTIIEFRGSLKVIIGASFDNDLSQLMLPDGLEKIESAFTGNP 360

QY 361 GDDHYNRRVVLWTYKSGKQPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420  
 DB 361 GDDHYNRRVVLWTYKSGKQPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

QY 421 GFSNKGLOKVKRNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480  
 DB 421 GFSNKGLOKVKRNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480

QY 481 KIGAFAPQSNNLKSFEASDDLBEIKEGAFMNNRIETLELKDCLVTIGDAAFINHIIYIV 540  
 DB 481 KIGAFAPQSNNLKSFEASDDLBEIKEGAFMNNRIETLELKDCLVTIGDAAFINHIIYIV 540

QY 541 LPESVQIEGRSAFRQNGANNLIIPMGSKVKTIGEMAFLSNRLEHLDLSEQQLTEIPVQAF 600  
 DB 541 LPESVQIEGRSAFRQNGANNLIIPMGSKVKTIGEMAFLSNRLEHLDLSEQQLTEIPVQAF 600

QY 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSHIAFNALDDNDGDEQFQNKVVVK 660  
 DB 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSHIAFNALDDNDGDEQFQNKVVVK 660

QY 661 THNSYALADGDEHPFVDPDKLSSITVDLEKILKLEGLDYSTLRQTTQTQFRDMTTAGKA 720  
 DB 661 THNSYALADGDEHPFVDPDKLSSITVDLEKILKLEGLDYSTLRQTTQTQFRDMTTAGKA 720

QY 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780  
 DB 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780

QY 781 AYNSALTKKANVKRLEKELDLTLGLVEGKGPLAQATWVGYYLLKTPLPPEYYIGLVNY 840  
 DB 781 AYNSALTKKANVKRLEKELDLTLGLVEGKGPLAQATWVGYYLLKTPLPPEYYIGLVNY 840

QY 841 FDKSGKLIYALDMSDTTIGEGQDAYGNPILNVDEDNEGYHALAVATLADYEGLDIKTILN 900  
 DB 841 FDKSGKLIYALDMSDTTIGEGQDAYGNPILNVDEDNEGYHALAVATLADYEGLDIKTILN 900

QY 901 SKLSQLTSSIRQVPTAAVHRAGIFQAIQNAAAEABQLLPKPGTHSEKSSSESANSKORGL 960  
 DB 901 SKLSQLTSSIRQVPTAAVHRAGIFQAIQNAAAEABQLLPKPGTHSEKSSSESANSKORGL 960

QY 961 QSNPKTNRGRHSALTPTGSGSPVYGILGTVTSVALLSLITAIKKKKY 1008  
 DB 961 QSNPKTNRGRHSALTPTGSGSPVYGILGTVTSVALLSLITAIKKKKY 1008

RESULT 4  
 AAU03612  
 ID AAU03612 standard; protein; 1055 AA.  
 XX AAU03612;  
 DT 12-SEP-2001 (first entry)  
 XX Group B Streptococcus antigenic protein, ID-87.  
 XX Group B Streptococcus; encapsulated bacterium; therapeutic; sepsis;  
 XX meningitis; neonate; antigenic; vaccine; infection; genital tract;  
 XX capsid polysaccharide vaccination.  
 XX Streptococcus agalactiae.  
 XX WO200132882-A2.  
 XX 10-MAY-2001.  
 XX 07-SEP-2000; 2000WO-GB003437.  
 XX 07-SEP-1999; 99GB-00021125.  
 XX (MICR-) MICROBIAL TECHNIQS LTD.  
 XX Le Page RWF, Wells JM, Hanniffy SB;



PS	Claim 18; Fig 2; 60pp; English.	
XX	The present sequence represents a BVH-A4 protein from serotype III Group B streptococcus (GBS) strain COH1 (designated GBS-BVH-A4) (I). (I) has antibiotic and immunostimulant activities, and can be used in polypeptide therapy and in vaccine production. (I) can be used for the therapeutic or prophylactic treatment of GBS bacterial infection in a host susceptible to GBS infection. In particular, (I) is useful for treating or preventing GBS infection in a neonate or infant (e.g. sepsis, meningitis, pneumonia, cellulitis, osteomyelitis, septic arthritis, endocarditis or epiglottitis), in a pregnant woman (e.g. mild urinary tract infection to life-threatening sepsis and meningitis, osteomyelitis, endocarditis, amniotia, endometritis, wound infection (post-caesarean or post-episiotomy), cellulitis or fasciitis), in a non-pregnant adult (e.g. bacteraemia, skin or soft tissue infection, pneumonia, urosepsis, endocarditis, peritonitis, meningitis or emphysema), or in a member of dairy herd (e.g. mastitis). A composition comprising (I) or (I) can also be useful for treating or preventing streptococcal infection	
XX	Sequence 1055 AA;	
SQ	Query Match 74.4%; Score 3777.5; DB 6; Length 1055; Best Local Similarity 71.7%; Pred. No. 1.7e-221; Matches 756; Conservative 105; Mismatches 140; Indels 53; Gaps 5;	
QY	2 KKHAKTVALTITVSVTHNQVSVLKVPEILKQTOASSISGADYAESGSKLKNET 61	721 LLSKSNLRQGEKQKFLQEAQPFGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780
DB	3 KKHAKTVALTITVSVTHNQVSVLKVPEILKQTOASSISGADYAESGSKLKNET 61	722 LLSKSNLRQGEKQKFLQEAQPFGRVDLDKAIKAEKALVTKKATKNGHLERSINKAVL 781
QY	62 SQPVDVTDLFSDKRTTPEKIKDNLAKGPRQELKAVTENT-ESEKQITSGSQLEQKE 120	781 AYNSAIKANKVRKLEKELDLITGLVSGKGLAQATWVGYYLLKTPPLPEYYIGLVNVY 840
DB	62 NSTVDETVDLFDGNSNNSSTESVSDPKQVPKAPKPEVTQEAASNSNDASKVEVPKQ 121	782 AYNSAIKANKVRKLEKELDLITGLVSGKGLAQATWVGYYLLKTPPLPEYYIGLVNVY 841
QY	121 SUSLNTVPSTNWEICDPTIKGNTLVLSKSGVEKLSQTDHLVLPQAAAGTQLIQVAS 180	841 FDKSGKLIYALDMSDTTGEQKDAYGNPILNVDNEDNGYHALAVATLADYEGLDIKTILN 900
DB	122 DTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLPFSAADGTQLTQVAS 181	842 FDKSGKLIYALDMSDTTGEQKDAYGNPILNVDNEDNGYHTLAVATLADYEGLYIKOILN 901
QY	181 FAPTPDKKTAIAEYTSRAGEEISQLDVDGKEIINEGEVENSLLKVVITPTGYKHGQ 240	901 SKLSQLSIRQVPTAAVHRAGIFQAIQNAEAEQQLPKPQTHSEKSSSSSANSKDRGL 960
DB	182 FAPTPDKKTAIAEYTSRAGEEISQLDVDGKEIINEGEVENSLLKVVITPTGYKHGQ 241	902 SSLDKIKAIRQIPLAKYHRLGIFQAIRNAEAEQQLPKPQTHSEKSSSSSANSKDRGL 960
QY	241 DAFVFNKNIAYNLPSLETISDYAPAHALAKQIDLPNLKAIAGELAFPDNQITGLSLP 300	961 QSNPK-----TN-----RGRHSA 973
DB	242 DAFVFNKNIAYNLPSLETISDYAPAHALAKQIDLPNLKAIAGELAFPDNQITGLSLP 301	958 EKNLKPVDYKTPIFNKALPNEKVDGDRAAKGHNINAEITNSNSVAVTPIRSEQLHKSQSDV 1017
QY	301 ROLMLAERAFKSNHKTIEFRGNSLKVIGEASPDNDLSQMLPDGLEKIESEAFPTGNP 360	974 ILPTGSGSVYGIYLTSTVALLSLITAIKKKK 1007
DB	302 RHLTKLAERAFKSNRIQTVFPLGSKLVIGEASPDNDLSQMLPDGLEKIESEAFPTGNP 361	1018 NLPQTSSKNNFYBILGVYSLCLLFLVTAGKKGK 1051
QY	361 GDHYNNVVLWTKSGKPSGLATENTYVNDPKSLWQESPEIDYTKMLEEDFTYQKNSVT 420	RESULT 6
DB	362 GDHYNNVVLWTKSGKPSGLATENTYVNDPKSLWQESPEIDYTKMLEEDFTYQKNSVT 421	ABP25812
QY	421 GFSNKGLOKVRKNKLEIPKQHNQVITIIEIGNAFRNVDFQNTLKYDLEBVKLPSTIR 480	ID ABP25812 standard; protein; 1055 AA.
DB	422 GFSNKGLOKVRKNKLEIPKQHNQVITIIEIGNAFRNVDFQNTLKYDLEBVKLPSTIR 481	XX ABP25812;
QY	481 KIGAFAPQSNNLKSPASDLEBEEKGAPMNNRIETLEKDKLVITIGDAAFHINHIYIV 540	XX 02-JUL-2002 (first entry)
DB	482 KIGAFAPQSNNLKSPASDLEBEEKGAPMNNRIETLEKDKLVITIGDAAFHINHIYIV 541	XX Streptococcus polypeptide SEQ ID NO 800.
QY	541 LPESVQETGRSFRONGANNILFMGSKVKTLCGEMAFSLNLEHLDSLSEKQLTEIPVOAF 600	XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
DB	542 LPESVQETGRSFRONGANNILFMGSKVKTLCGEMAFSLNLEHLDSLSEKQLTEIPVOAF 601	XX group A streptococcus; Streptococcus pyogenes; antibacterial;
QY	601 SDNALKEVLLPASLTIRREAFKQNLKQLEVASALSHIAFNALDNDGDEQFNKVVVK 660	XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
DB	602 SDNALKEVLLPASLTIRREAFKQNLKQLEVASALSHIAFNALDNDGDEQFNKVVVK 661	XX Streptococcus agalactiae..
QY	661 THNNSYALADGHHFTVDDPKLSSTIVLEKILKLEGLDYSLRTOTTQFRDMWTAGKA 720	XX WO200234771-A2.
DB	662 THNNSHMLADGRRFTIDPDKLSSTVMDLEKVLKIIIEGLDYSLRTOTTQFRDMWTAGKA 721	XX 02-MAY-2002.
		XX 29-OCT-2001; 2001WO-GB004789.
		XX 27-OCT-2000; 2000GB-00026333.
		XX 24-NOV-2000; 2000GB-00028727.
		XX 07-MAR-2001; 2001GB-00005640.
		XX (CHIR-) CHIRON SPA.
		XX (GENO-) INST GENOMIC RES.
		XX Telford J. Masignani V, Margarit Y Roel, Grandi G, Fraser C;
		XX Tettelin H;
		XX WPI; 2002-352536/38.
		XX N-PSDB; ABN66443.
		XX New Streptococcus protein for the treatment or prevention of infection or
		XX disease caused by Streptococcus bacteria, such as meningitis, and for
		XX detecting a compound that binds to the protein.
		XX Claim 1; Page 3230-3231; 4525pp; English.
		XX The invention relates to a protein (ABP25413-ABP30895) from group B
		XX streptococcus (Streptococcus agalactiae) or group A streptococcus (GAS
		XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
		XX the specification. The proteins have antibacterial and antiinflammatory
		XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
		XX antibodies that bind (I) are used in the manufacture of medicaments for
		XX the treatment or prevention of infection or disease caused by

Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.  
Nucleic acids encoding (I) are used to detect *Streptococcus* in a  
biological sample. (I) is used to determine whether a compound binds to  
(I). A composition comprising (I) or a nucleic acid encoding (I), may be  
used as a vaccine or diagnostic composition. The disease caused by  
*Streptococcus* that is prevented or treated may be meningitis. Nucleic  
acid encoding (I) may be used to recombinantly produce (I) and may be  
used in gene therapy. Antibodies to (I) are used for affinity  
chromatography, immunoassays, and distinguishing/identifying  
*Streptococcus* proteins  
Sequence 1055 AA;  
XX  
SQ

Query Match	74.3%;	Score 3776.5;	DB 5;	Length 1055;
Best Local Similarity	71.7%;	Pred. No. 1.9e-221;		
Matches 756;	Conservative 105;	Mismatches 140;	Indels 53;	Gaps 5

Qy	2	KGHLKTVALTITTVSVTHNOEVSFLVKEPILKOTQASSISGADYAEBSGSGKSKINET	61
Db	3	KGHLKTLALALTTVSVVTVSQEYGLERESVKEQTQSA-SEDDWFERDNERKTWVSKE	61
Qy	62	SGPVDVTDLFDSDKRTTPEKIKONLAKGPREQELKAVTENT-ESEKQITSGSOLEPSKE	120
Db	62	NSVDETVDLFDSDGNSNNSSKSTESVSDPQVPKAKPEVTQBSNNSNDASKVEVPKQ	121
Qy	121	SLSLNKTVPSSTNWEICDPIITKGNLTNLVGLSKSGVEKLSQTDHLVLPSQAADGTQLIQVAS	180
Db	122	DTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLPSHAADGTQLTQVAS	181
Qy	181	FAFTPDKKTAAIABYTSRAGENGEISOLDVDGKEIINEGEVFNVSYLKKVTIPTYGKHICQ	240
Db	182	FAFTPDKKTAAIABYTSRLENGKPSRLDIDQKEIIDEGEIFNAYQTLKTIPTYGKISIQ	241
Qy	241	DAFVDMKNIAEYNLPESLETTSDYAFAHALKAQIDLPNLKAIGELAFDFNQITGKLSLP	300
Db	242	DAFVDMKNIAEYNLPESLETTSDYAFAHMSLKQVKLPDNLKVIGELAFDFNQIIGGKLILP	301
Qy	301	ROMLRAERAFKSNHIKTIEPRGNSLKVIGEASQDNDLSQMLPDLGLEKIKISEAFTGNP	360
Db	302	RHLIKLAERAFKSNRIQTVEFLGSKLVIGEASQDNNLRNVMLPDLGLEKIKISEAFTGNP	361
Qy	361	GDDHYNNRVLWTKSGKNPSGLATENTVYNPDKLSLWQESPEIDYTKWLEEDFTYQKNSVT	420
Db	362	GDBHYNNQVVLKRTTQGMPHQLATENTVYNPDKSLMRATPDMQDITWLEEDFTYQKNSVT	421
Qy	421	GFSGNLGQVKENKNLEIPKOHNGVYTIITEIGDNAFRNVDPQNTLARKYDILEEVLKLPSTIR	480
Db	422	GFSGNLGQVKRENKNLEIPKOHNGITITEIGDNAFRNVDPQSKTLRKYDYLEEKLPSTIR	481
Qy	481	KIDAFAPQSNLKSFEASDDLEEKEGAFPMNRIETLELKKDLVTTGDAAAFHNHIIAYIV	540
Db	482	KIDAFAPQSNLKSFEASEDLEETKEGAFPMNRIETLIDLKDKLIKIDGAAAFHNHIIAYIV	541
Qy	541	LPBSVQEIIGRSFRONGANNILFMGSKVKTILGEMAFLSNRLHLDSLSEOKOLTEIPVQAF	600
Db	542	LPBSVQEIIGRSFRONGALHLMFIGNVKVKTILGEMAFLSNKLSEVNLSSEQKOLTEIVQAF	601
Qy	601	SDNALKEVLLPASLTKTIREEAFKQNHUKOLEVASALSHIAFNALDNDGDGEQDNKVVVK	660
Db	602	SDNALSEVLLPNLQTIREEAFKRNHLKEVKGSSLTLSQITFNAFDQNDGDKRFGKVVVR	661
Qy	661	THHNSVALADGPHFVDPDKLSSTIVOLEKTLKLIIEGLDYSTLROTTQTOFRDMTTAGKA	720
Db	662	THHNSHMLADGGRFIIIDPDKLSSTVMDLEKVLKIIIEGLDYSTLURQTQTOFREMITTAGKA	721
Qy	721	LLSKSNLROGEKQKFLQEAQFPLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL	780
Db	722	LLSKSNLROGEKQKFLQEAQFPLGRVDLDKAIKAEKALVTKKATKNGHLLERSINKAVL	781
Qy	781	AYNNSAIKANVRILEKELDLTLTGLVEKGPLQAOATWVGQVYLLKTPPLPYEYIGLANY	840
Db	782	AYNNSAIKANVRILEKELDLTLTGLVEKGPLQAOATWVGQVYLLKTPPLPYEYIGLANY	841

RESULT 7  
ADH62807  
ID ADH6

ID ADH62807 standard; protein; 4734 AA.

AC ADH62807:

XX DT 15-APR-2004 (first entry)

5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528

XX mucin binding; probiotic; glycoprotein; epithelial cell;  
KW gastrointestinal tract; lung; uterine cervix.

*Lactobacillus johnsonii*.

XX  
DN  
FD1382970-A1

XX 21 JAN 2004

XX

•

**X**

XX  
. (NES1 / 300

PI PR1dmore RD;  
YY

DR WPI; 2004-111519/12.

XX Novel mucin binding polypeptide encoded by novel gene of *Lactobacillus*  
PT *johnsonii*, useful for binding bacteria to mucin.  
PT

PS Claim 8: SEO ID NO 25: 225pp: English.

This invention relates to novel mucin binding polypeptides and the DNA sequences which encode them. In particular, the genes are derived from *Lactobacillus johnsonii*. The invention may allow a better understanding of the mechanism by which probiotics (such as *L. johnsonii*) can bind to mucins, large glycoproteins present on the surface of a large number of epithelial cells including those in the gastrointestinal tract, the lung or the uterine cervix. The present sequence is that of an *L. johnsonii* mucin binding protein of the invention.

XX  
SC  
Sequence 4734 AA.

Query Match	4.8%;	Score 243;	DB 8;	Length 4734;
Best Local Similarity	19.0%;	Pred. No. 5.1e-05;		

QY	21	NOEVFSLVKEPILKOTQASSISGA	DYAESSG	KSKLKINETS - 62
		. . . . .		

Qy	63	--GPVDTVTDLFSDKXRTPEK---IKDNL-AKGPREQEL-----KAVTENTES 105
Db	3680	DEEKAALKQEVSDAQTAANTADNATTNAAVTEAEDNGIKAINSLSEVPTKSDAKEQATSD 3738

Db 3740 LNSAVDEAKAIDQSDNLTDDEBKQVAKAQIDSDAKKAQEAIDTAKTNDVDVKKAIIDGTILA 3799  
QY 106 -EKQI-----TSGSQLEQSK-----ESLSANKTYPSTSNWEICDPITKNTLVGLS 150  
Db 3800 IDKDVANNAIDNAVAGKKAEBISKPLTDEEKTALNNEVDEKAN-SAKDAINKATTPEGVT 3858  
QY 151 KSGVEKLSQTDHLVLPQ-----AADGTQLIQVASFAP---TPDK 187  
Db 3859 EAQSGIKSIDDVNPTTSVAKAEAKKAVAEAAEKNAIDSSNLTDEKALKQAEVSDA 3918  
QY 188 KTAIAEYTSRAGEGEISQLDVDGKEIINEGEVFNYSYLLKKVT-----IPTGYKHIGQ 240  
Db 3919 QTAANTAIDNAITNAEVAEDNGVKTINGIEVPTKSTTKEQATNDLNEVENAKAIDQ 3978  
QY 241 DAFV--DNKNIAEVLNPSLETISDYAFALHAKQIDLPDNL-KAI--GELAF----- 288  
Db 3979 DSNLTDEEQVAKQID-----SADKAQADAINNAKTNDVDVKKAIIDGTTLAIDKDVANA 4032  
QY 289 -PDNOLTGKLSPLRQMLRAERAFKSNHIKTIEPRGNSLKVIGEASFQDNDLSQLMLPDG 347  
Db 4033 AIDNAVAGKLEINSNLNEEKQAVTDLIN-----NEADNAKQIAEATTPEE 4080  
QY 348 LEKIESEFTGNPGDDHNNRVLMWTKSGKNPSGLATENTYVNDKSLMQBESPEIDYTKW 407  
Db 4081 VTRAQEE-----GVKNININVPPTSPAKDAANAAL-----DQALKNKKDEINNATN 4127  
QY 408 L-----EEDFTYQNSVTFGSNGLOKVKGNKLEIPKQHNGVTIETI-----GDNAFRNV 458  
Db 4128 ISSEKTLIKQATEAANIAKONINNATTNSEVETAQVDGEKAIADVTGPGLSIDIKKESI 4187  
QY 459 DPQNTLRKYDUEVKLPSTIRKIGAFAPQSFQSNLKSFEASDLEIEKGAFF-MNNRIETL 517  
Db 4188 DLINKALNEKQ-DEIN-----NANLSQDESTEIDIOAKKIATETAINNEINA 4233  
QY 518 ELKDKLVITIGDAFRHNIYAIVLPSQVEIGRSA-----FRQGANNLIPMGSKV 568  
Db 4234 QTNDEAKAADTG--VKNIENVISIP-SIEDACKNATQAIIDDALNSKKNNEINNASLITDSE 4290  
QY 569 KTLGEMAFLSNRLEHLDLSEKQLTEIPVQAFSDNALKEVLPLPASLKTIRBEAFKKNHLK 628  
Db 4291 KT-----DLI--NQATEIA-----NAAKDAINSATNTAVEAAEYKG--- 4325  
QY 629 QLEVASALSHIAFNALDNDGDGEQPDNKKVVKVTHNSVALADGEHFIVDPDKLSSTIVDL 688  
Db 4326 -----VADINNIHFTNLDSS-----KKAANS-AIEDA----- 4351  
QY 689 EKILKLIBGLDYSLRTQTOTQFRDMTTAGKALLSKNSLRQGEKQFLQEAQFGLGRVDL 748  
Db 4352 -----LTTKKDEINNASNLSDSEKAKLINQA----- 4377  
QY 749 DKAIKAERKALVTKATKNGQLLE-----RSINKAVLAYNNSA-----IKKANV 792  
Db 4378 -TEIANAAKAAI--NNATNSAVTAENKGIEDIANINVPSLAETKQAIDAIDAIQVQKAKN 4435  
QY 793 KELEKELDILT---GLVEGKPLAQATWQGVYLLKTPPLPEYVIGLVNVPFDSKGLI 848  
Db 4436 SQIEEAKNSLADSEQKXNLDQVNNKIAQ-----DAINKLNDPATTTN-----EVITDTRKAI 4486  
QY 849 YALD--MSDTIGEGQKDAY-----GPN 868  
Db 4487 DQITNLFIPTLDSVQKDAQEAINSAQETKIDINKADNLTDQMKQNLIDQVDQVADKATK 4546  
QY 869 INLVDEDEGHYHALAVATLADYEGLDIDKITLNSK----- 902  
Db 4547 AINNAQTNDVDVKEABIEGLEIDTSIKVPSLVESKDDAIKEINDALKKKKTDEINNAADLDQK 4606  
QY 903 -----LSQLTSTRQVPTAAYHRAGIQAIONAAEAE-----QLLPKCTHSEKSSS 949  
Db 4607 QKDELIQSITDI-----ATETKTKVFNATTNAEVDAAEAEAGIKAEVKIPARTADNNT 4661  
QY 950 SESANSKDRGLQSN---PRTNRGRH-----SAILPRTGSKGSFVVGILGYTSVALLS 998  
Db 4662 --ESESKEQTVITNSVQPKRNAVHKNGTPVNNKATLPQTGCKKNLTITLAGAALLGLAG 4719

QY 999 LITAI-----KKKK 1007  
Db 4720 VFSLFGLGDKRKK 4732

## RESULT 8

ADS22318  
ID ADS22318 standard; protein; 669 AA.

XX AC ADS22318;

XX DT 02-DEC-2004 (first entry)

XX DE Bacterial polypeptide #11351.

XX KW Recombinant DNA construct; transformed plant; improved plant property; cold tolerance; heat tolerance; drought tolerance; herbicide; oomosis; pathogen tolerance; pest tolerance; plant disease resistance; cell cycle pathway modification; plant growth regulator; homologous recombination; seed oil yield; protein yield; carbohydrate; nitrogen; phosphorus; photosynthesis; lignin; galactomannan; bacterial polypeptide.

XX OS Bacteria.

XX PN US2003233675-A1.

XX PD 18-DEC-2003.

XX PF 20-FEB-2003; 2003US-00369493.

XX PR 21-FEB-2002; 2002US-0360039P.

XX PA (CAOY/) CAO Y.

PA (HINK/) HINKLE G J.

PA (SLAT/) SLATER S C.

PA (CHEN/) CHEN X.

PA (GOLD/) GOLDMAN B S.

PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

WPI, 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.

PS Claim 1; SEQ ID NO 11351; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition. This sequence represents a bacterial polypeptide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic





Db 355 GSSAFQGCISITSVNIGSGVTSIGNSAFQCSALTTVTIPDSVTTILDYAFQDCSRLTTV 414  
 QY 347 ----GLEKIESAFTCNPGDDHNNRVLTWTKSGNPSGLATENTYVNPDKSLWQESPEI 402  
 Db 415 TIGSGVTNIDSAFSG-----CSALTAID--VDNDSIYASIDGV 452  
 QY 403 DYTCKLEEDFTYQKNSVTFGSKGLQKVKRNKNLEIPKQHGNGVITETGDNAFRVVD-- 459  
 Db 453 LYNK-----DITTLIQCPG-----KTSVPIPD-----SVTTIGDSAFRDCSGLT 493  
 QY 460 -----FQNKTLKRYDLEEVKLPSTIRKIGAFAPQS--NNLKSFEASDDLEE 503  
 Db 494 IVPIGNNVTSIGSSAFYGCY----LTSVPIGNNVTSIGSSAFYGCYTSITVIGNNVTS 549  
 QY 504 IKEGAFMN-NRIETLELKDVLITIGDAAPH-----INHIYA 538  
 Db 550 ISSAFQCSALTTVTIPDSVTTIGDIAFSDCSGLTTVTIGNNVTSIGSSAFYGCYCTALT 609  
 QY 539 IVPESVQIGRSAFRQNGANNLIPMGSKVKTGEMAFLS 578  
 Db 610 IAPDSVTSIGSADFRCNTLTSVTIGNNVTSIGSSAFRS 649

## RESULT 10

ABB67502  
 ID ABB67502 standard; protein; 3257 AA.

XX ABB67502;

AC ABB67502;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 29298.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-658660/75.

DR N-P8DB; ABL11605.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

XX Disclosure; SEQ ID NO 29298; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB27072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 3257 AA;

Query Match 4.3%; Score 220.5; DB 4; Length 3257;  
 Best Local Similarity 18.9%; Pred. No. 0.00071;  
 Matches 221; Conservative 185; Mismatches 437; Indels 329; Gaps 51;  
 QY 2 KXHLKTVALTTLTVSVVTH-----NQBFSLVKEPI--LKQTOAS-----SSISGADYAES 50  
 Db 1206 KKEAVVGPDLKDTSTSSNIIDKKSNEFSFDSAMQPSDLNQLKQESAFTKLSISSPKKIMK 1265  
 QY 51 SGKSKLKNETSQPVDDTVTLFSDKRTTPKIKDNLAQPRE--QELKATVENTESEKQ 108  
 Db 1266 DQDKDLDALSKGDSNPRTIRDTGDSRQTDKQHQENTKHEERDSKLKANIDETKSSSE 1325  
 QY 109 -----ITSGSQLEQSKESLSLNTKVPSTSNWEICDFITKGNLTVGLSKSGVEKLSQTDHL 163  
 Db 1326 KDAEPIKSDSQDSAKPRLSKPKSRNKKNE-----KKPND--SIAESDIEGGFQVN-- 1376  
 QY 164 VLPQAAADGTOLIQVASFAFTP--DKTAIAEYTSRAGENGESQLDVGDKKEIINGEV 220  
 Db 1377 -----TETVQAT--CSTPESNKKDMVK--SDETNEBPNLSETEIG--RIRKRGQA 1421  
 QY 221 FNSYLLKKVITPTGYKHIGQDAFVDNKNIAEAVNPESLETISDYAFALHAKQIDLPNL 280  
 Db 1422 FH-----IENPKDDLHITPON--ENOSIAGVNF-----KQVLPESV 1457  
 QY 281 KAIGELAFFDNOITGKLSLPRQLMLAERAFKSNHIKTIBPRGNSLKVIGEASFQDNDLS 340  
 Db 1458 ES-----DTPI--MKIPTKTYLMCTK-----NKTSL--SASEDDPI-- 1490  
 QY 341 QLMPLDGLKIESBAFTGNPGDDHNNRVVLTWTKSGRNP----- 379  
 Db 1491 -VLSPQKL--ITTSKGSNPDLDNANN--LETSTQDPKEHEFSQDTFTDNSDIIPST 1544  
 QY 380 -----SGLATENTYVNPDKSLWQESPEIDYTKMLEDFYQKNSVTGFSKNG 426  
 Db 1545 KKSQIVPPTPTKSSDQTKNSFITNRSKPSKRVNSKEAKLDNSFEESQNAEASSASK 1604  
 QY 427 LQK-----VKENKNLEIPKQHGNGVITTEIGDNAPRVDFONKTLKRYDLE 471  
 Db 1605 VQKELRPTTASCRKLRVLIKRTPTSSLP-----TNSRKSIFKTPAKSKRLTKILES 1656  
 QY 472 EVKLPS-----TIRKIGAPA-----FQSNLKSFEASDDLEIKEGAF 509  
 Db 1657 MEKTPSREPSVSLGEVNPDSPPVAASVAVLHESDRDLESNEIPNEEVFDETE--ASAED 1715  
 QY 510 MNNRIETLELKDVLITIGDAAFHINHIYAIVLPESVQIEGRSAFRQNGANNLIPMGSKVK 569  
 Db 1716 TDNKLKKKE-----DDHELVNDICAASKNPITDDTKDASSNKTSDSVLQETKDE 1767  
 QY 570 TLGEMAFLSNLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREBAFKKNHLKQ 629  
 Db 1768 -----LSNSLINATQGEDTPIKELTEEEVPNN-----KTVEDESKQEIILKD 1809  
 QY 630 LEVASALSHIAFNALDDNDGD-----EQFQDNKVVVTKTHNSYALADGHEHFTVDPDKLS-S 683  
 Db 1810 LEPDNA-----ALEEDTASTAKAEEMDLVIKEKSNVKS-VLAEPETDVTDDLELAQS 1861  
 QY 684 TIVLEKILKILGLDYST---LRQTTQTQFRDMTTAGKALLSKSNLRQ-----GSKQ 733  
 Db 1862 PIPNSSETTSDTDPPEPSTSSVWKRSLRKREADSSQDPEA--AKRKQKQDVEKSLTGKKE 1919  
 QY 734 KFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGOLLERSINKAVLAYNNNSAIKKAN-- 791  
 Db 1920 QVKPARRQLAEVE--ERPSLKRSTESSEKSTVQGYKISITIGNETIMSSSTAPIRENRE 1978  
 QY 792 -----VKRLKEKELDLITG-----LVEGKGPLQAQATMVQGVYLL 824  
 Db 1979 AASTSPSARKSAVQEAQKHVETTKHILGPPCKILLHSDSPAEEKPMVQ--TLLSSTLSL 2037  
 QY 825 KTPILPLPYIYGLNVYFDKSGKLIYALDMSDTIGEGQKDAYGNFILNVDENEGHALAV 884  
 Db 2038 QKPSTL-----DDGSPLKIRKSLKKSIAD-----NIDGQSQSFSSSV 2076  
 QY 885 ATLADYEGLDIKTILNSKLSQITS--IRQVPTAAVHRAGIFQAIQNAEAEQQLLPKPGTH 943



Db 2077 --LNKNTSVVAPRKVNISVLLQSKDTQVETAA-----SSSETPLTKKEXLK 2122  
 QY 944 SEKSSSESAN-----SKRGLOSNP--KTNRGHSAAILPR-----TSGSGSFYVGI 988  
 Db 2123 TQSTKKGPNKTESKSKSLVQGPOMKQKSEEAIVGPKILNKYLKSETESSRKTSTV 2182  
 QY 989 LGYTSVALL-----SLTAIKKKK 1007  
 Db 2183 TGRKIQIGLEVLKPKESRKSSESLVEAISRK 2214

## RESULT 11

ID ABB54963

XX ABB54963 standard; protein; 1072 AA.

AC ABB54963;

XX 29-AUG-2003 (revised)

DT 16-MAY-2002 (first entry)

XX Lactococcus lactis protein yqfG.

XX Lactococcus lactis protein yqfG.

XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

XX Lactococcus lactis; IL1403.

OS Lactococcus lactis; IL1403.

PN FR2807446-A1.

XX 12-OCT-2001.

XX 11-APR-2000; 2000FR-00004630.

XX 11-APR-2000; 2000FR-00004630.

XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.

XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;

XX WPI; 2002-043418/06.

XX New nucleotide sequence useful in the identification or Lactococcus

PT lactic and related species.

XX Claim 6; SEQ ID NO 1665; 2504pp; French.

XX The present invention is related to a Lactococcus lactis nucleotide

CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic

CC acid sequence is useful in the detection and/or amplification of nucleic

CC acid sequence, particularly to identify Lactococcus lactis or related

CC species. The proteins of the invention are useful for the biosynthesis or

CC biodegradation of a composition of interest. The invention helps research

CC in lactic bacteria, particularly useful in the production of yogurt and

CC cheese. Note: The sequence data for this patent is based on equivalent

CC patent WO200177334 (published 18-OCT-2001) which is available in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 29-AUG-2003 to

CC standardise OS field)

XX Sequence 1072 AA;

SQ

Query Match

Best Local Similarity 4.3%; Score 217.5; DB 5; Length 1072;

Matches 151; Conservative 159; Mismatches 365; Indels 219; Gaps 29;

QY 149 LSKSGVEKLQSOT-----DHLVLPQAADGT-----QLIQVASFAFTPKKT 189

Db 1 MSQSIKKITMTVGVLTLATLNPFLNQLITELPVEATTSTKITVTDEQKITYNLDKVS 60

QY 190 AIAEYTSRAGENG-RISQLDVQDKIINEGE-----VFNYSLLKKVTIPTGYKHI 238

Db 61 LTAEVASYAGLAGADIVIPDI-----IVNQGTYAITSIGTYAFNSGIRSVIIGNNVVDI 116

QY 239 QGDAFVD-----NKNIAEVNIPESLETISDYAFALHAKLQIDLPNKAIGELAFPD 290  
 Db 117 NTSAQTTTAPDYKXSLTKVILGAKVQNIKTDAFAGNAISSIEFPNSVLKIATRAPAN 176  
 QY 291 NOITGKLSLPQMLRAERAPKSHIETIEPRGNSLKVIGASPODNDLSQMLPDGLEK 350  
 Db 177 NNLT-ELSLGNITEIMAKAFQSNQITTEFADESLLTVDSAAFSGSSVQSLLTIGV-- 233  
 QY 351 IESEAFNGPCGDDHYNNRVVLWTKSGK-----NPSGLATENTYVNPDKSLWQES 399  
 Db 234 -----TLADDVFNKTSPLFGQUSDLPTDEIRTISVNSSGL-FDKSWINDSS--QS 281  
 QY 400 PEIDYTKWLEEDFTYQKNSVTGFSNGKLGQVKRNKNLEIPKOHNGVTI-----TEIGDN 453  
 Db 282 TSVS-----TENADTTVPSSSSNVEQVDVASS-ESTQDANSASLYPISEASVTDN 331  
 QY 454 AFRVDFONKTLRKVDLEEVKLPSTIRKIGAFQSNLKLKFEASD-----DLEE 503  
 Db 332 TLNSISLSDSSISSQSTENSQ--SGASSTABISYDSSENSLSLSSNQINSNSEKDSNQ 389  
 QY 504 IKEGAFNNRIETLELKDVLTVIGDAAPHIHVAIVLP---ESVQEIGRGAFRONGANN 560  
 Db 390 SSLGSSWSNESEHSNSNINETNNSEIINLLPSPNPTESNSVSDQTSSEASTNSNS 449  
 QY 561 LIFMGSKVKTILGEMAFLSNRLEHLDLSEKQLTEIPVOAFSDNALKEVLLPASLKTIREE 620  
 Db 450 ISLSPSNISSTSDSESATNSDFSNVAEVANNSLASVNNSSSVLSSTSTADNL-GINQS 508  
 QY 621 AFKNHLKQLEVASALSHIAFNALDDNDGDFQFNKVVVTKHNSYALADGEHFTVDPK 680  
 Db 509 GSDNLTKDSSEISTSGAFLSSNQTSE-----ASTNSNS-----SLSLSPSN 550  
 QY 681 LSSSTIVLEKLKLEGLDYSTLQTTQTQFRDMMTAGKALLSKNLKQGEKQKFLQEAQ 740  
 Db 551 ISSTSV-----LESTTSSNFSNVAEVANNSLASVNNSSSVLSSTSTADN-----LEINQ 601  
 QY 741 FFLGRVDLDKAIKAEKALVTKATKN-----GQLLERSINKAVLAYNNSAIKKANVK 793  
 Db 602 FGSDNLTKDSSEISTSGAFLSSNQTSEASNSMSSINSPSLSLSLTSSESATNQS-- 659  
 QY 794 RLEKELDLTLGLVEGKPLAQTWVGVLKTLPLPEYVIGLVNVPYFKSGKLIYALDM 853  
 Db 660 -----SSEATKV-----DNNNS-----STHS 674  
 QY 854 SDTIGEGOKDAYGNPILNVEDNEGYHALAVATLADYEGLDIKTILNSKLSQLTSTIRQVP 913  
 Db 675 SNILNSGNSDSDS---DSDSDSDS-----SNLSSSPNLETNQTITSSKPSSEVNNISNP 725  
 QY 914 TAAVHRAGIFQAIQVAAAEARQLLPKPGTHSEKSSSESANSKORGLQSNPKTN 967  
 Db 726 KKV-----SSSNSVQENSTDHEMSTNPKSS 750  
 RESULT 12  
 ID ADK47748  
 XX ADK47748 standard; protein; 641 AA.  
 AC ADK47748;  
 XX 20-MAY-2004 (first entry)  
 DT Streptococcus pneumoniae protein, Seq ID No 4263.  
 XX Streptococcus pneumoniae protein, Seq ID No 4263.  
 KW Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.  
 XX Streptococcus pneumoniae.  
 OS Streptococcus pneumoniae.  
 PN US6699703-B1.  
 XX 02-MAR-2004.  
 PD 26-MAY-2000; 2000US-00583110.  
 PF  
 XX

PR 02-JUL-1997; 97US-0051553P.  
PR 12-MAY-1998; 98US-0085131P.  
PR 30-JUN-1998; 98US-00107433.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;  
XX  
DR WPI; 2004-212399/20.  
DR N-PSDB; ADK45087.  
XX  
XX New nucleic acid molecules and polypeptides useful for diagnosing,  
PT preventing and treating pathological conditions resulting from bacterial  
PT infection, e.g. Streptococcus pneumoniae infection, and in drug  
PT screening.  
XX  
XX Disclosure; SEQ ID NO 4263; 301pp; English.  
XX  
XX The invention relates to isolated Streptococcus pneumoniae nucleic acids  
CC and polypeptides. The nucleic acids and proteins are useful for  
CC diagnosing, preventing and treating pathological conditions resulting  
CC from bacterial infection, such as S. pneumoniae infection. These may also  
CC be used for drug screening procedures. The present sequence represents a  
CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence  
CC data for this patent did not appear in the printed specification but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html.  
XX  
XX Sequence 641 AA;  
SQ  
Query Match 4.2%; Score 215.5; DB 8; Length 641;  
Best Local Similarity 23.2%; Pred. No. 0.00015;  
Matches 132; Conservative 72; Mismatches 178; Indels 187; Gaps 30;  
QY 6 KTVALTTLTVSVV-----THNQEVFS-----LVKEPIILKQ 35  
DB 3 KTTILSLTAAVILAAVYPNEPILADTPSEVIVKTKVGSIIQQNNIKYKVLTVEGNIRT 62  
QY 36 TQASSISGADY-ARSSGK-----SKLKINETSGPVDVTDVTLFS-----DK 76  
DB 63 VQGVNGVTPVEAGQDGKPFIPTKITVGDVFTVTEVASQAFSVYDPTGRIVYPPSS 122  
QY 77 RTTPEKIDNLAGPREQELKAVTENTSEKQITSGSOLEQSKB-----120  
DB 123 ITIPSSIKKIQQKGFHGSKAKTII-----FDKGSQLEKTEDRAFDFSELEIELPAS 174  
QY 121 -----SLSLNK-TVPSTSNWEICDFTKGNLTGLVLSKSGVEKLSQTDHLVLPSSQ 168  
DB 175 LEVITGSAPSPQKLKLTFFSSSKLEL-----ISHEAPANLSNLEKLTLPKS 222  
QY 169 AAD-OTQLIQA-----SPA-----PTPKKTAIAEYTSRAGENGESIQD 208  
DB 223 VKTLGSLNFRLLTSLKHVDVERGNSFASVDGVLPFSK-DTQLIYPPSQ--KNDESXYTP 279  
QY 209 VDGKIIINEGEVFNFSVLLKKVTIPTGYKHIGODAFVDNKNIAEVNLPESLETISDYAF-A 267  
DB 280 KETKELASVFNKNSY-LKKLSNLEGLEKIGTFAPADAIFAUEISLPSNLETIERLAFYG 338  
QY 268 HIALKQIDLPDLKKAIGELAF-----PDNQTGKLSLPQMLRLAERAFKSNHK 317  
DB 339 NLELXELILPDNVKVPFGKHVWNGLPKLSLTTGNIN---SLPSPFLSGVLDLSEIHK 395  
QY 318 -----TIEFRGNS--LKVIGEASFQ-DNDLSQLMLPDGLEKIESRAPT 357  
DB 396 NKSTEFVSKDTFAIPETVKFYVTEHINKDLKLSNLSNDI-----IVEKVDNIQETDV 451  
QY 358 GNPGDHYNRNVVLWTK-----SGKNPSGLATENTYVNPDKSLW---QESPEIDY 404  
DB 452 AKP-KKNSQGVGVGVKQKGLWYLYNLSGSMATGV-----KQKGLWYLYNLSGSM-A 502  
QY 405 TKWLEED---FTYQKNS---VTGF-SNKGKL 427  
DB 503 TGWVKDKGLWYLYNLSGSMATGVKDKGL 531

RESULT 13  
ADR94313  
ID ADR94313 standard; protein; 641 AA.  
XX  
AC ADR94313;  
XX  
DT 16-DEC-2004 (first entry)  
XX  
DE Novel S. pneumoniae protein sequence, SEQ ID 2948.  
XX  
KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
KW bacterial infection.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN US6800744-B1.  
XX  
PD 05-OCT-2004.  
XX  
PF 30-JUN-1998; 98US-00107433.  
XX  
PR 02-JUL-1997; 97US-0051553P.  
PR 12-MAY-1998; 98US-0085131P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
DR WPI; 2004-697205/68.  
DR N-PSDB; ADR91710.  
XX  
PT New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 2948; 151pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR84489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridisable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acid and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed S.  
CC pneumoniae protein sequences. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.  
XX  
SQ Sequence 641 AA;  
Query Match 4.2%; Score 215.5; DB 8; Length 641;  
Best Local Similarity 23.2%; Pred. No. 0.00015;  
Matches 132; Conservative 72; Mismatches 178; Indels 187; Gaps 30;  
QY 6 KTVALTTLTVSVV-----THNQEVFS-----LVKEPIILKQ 35  
DB 3 KTTILSLTAAVILAAVYPNEPILADTPSEVIVKTKVGSIIQQNNIKYKVLTVEGNIRT 62



Db 700 NNSAFTALSEBDQLL---SQVKELSMVTELRAQVKQLENNLAEEAQRRLDYESTAH 756  
QY 602 DNALKEVL-----LPASLKTIRREAFKONHLKQLEVASALSHIAFNALDNDGDGDFDNK- 656  
Db 757 DNLLTEQIHSLISIRAKSKDVKIEVL-QNELDDVQLQFSEQSTLIRSL-----QSOLQNK 810  
QY 657 -----VVVKTTHNSVALADGHHFIVDPKLSSTIVDLKSLIKLIEGLDYSTLR 704  
Db 811 SEVLGAEVRHRSKVBELSQALSQKLEIITQMOLL-----LEK-----KRDVETLQ 859  
QY 705 QTOTQFDMITAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKA 764  
Db 860 QTIEBKDQV-----EISFSTKRWQLNEEKFSIG-VEI-KTLKEQLNLSRAEE 909  
QY 765 TNGQLLERSINKAVLAVNNSAIKVA---NVRLKELDLTLGLVEGKPLAQATWQGV 821  
Db 910 AKKEQVEDNEVSSGLKQNYDMSAPGQTSKEELQHEFDLLKENEQRKRLQALINRK 969  
QY 822 YLLKTPPLPEYIYGLNVYFDSKGLIVALDMSDT-IGSGOKDAYGNPILNVDEDEGHH 880  
Db 970 ELLQVRSLREELANK---DESKK---EIPUSETERGEVEE-----DKNKEYS 1013  
QY 881 ALAVATLADYEGLDIKTILNSKLSQTSIR-----QVPTAAVHRAGI 922  
Db 1014 EKVTSKQCEIEIYIKQTSISEKEVELQHIRKDLBEKLAEEQFQALVKQMOTLQDKTNQ 1073  
QY 923 FOAIONAAAEQLPKPGTGHSEKSSSES-ANSKDRGLQSNPKTNRGH 971  
Db 1074 IDLLQAEISENQAIQKLITNTDASDGSVALVKETVVISPPCTGSGSEH 1123

## RESULT 15

ADSE5037  
ID ADE56037 standard; protein; 3259 AA.  
XX  
AC ADE56037;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human Protein CAA53052, SEQ ID NO 1876.  
XX  
KW Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
OS Homo sapiens.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GENO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
XX Woolf C, D'urso D, Befort K, Costigan M;  
XX WPI; 2003-268312/26.  
DR GENBANK; CAA53052.  
XX  
XX New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
XX Claim 1; Page; 1017pp; English.  
XX  
XX The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also

CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3259 AA;

Query Match 4.2%; Score 214.5; DB 7; Length 3259;  
Best Local Similarity 19.8%; Pred. No. 0.0016;  
Matches 212; Conservative 168; Mismatches 427; Indels 263; Gaps 45;  
QY 20 HNEVPSLVKPEILKQTOASSISGADYAESGSGSKLKI-----NETSPVDVTVDL 72  
Db 233 HEDELLQIV-----TQAD-----VETEMQKLRVLQKLEHEESLVGRAQVVDL 277  
QY 73 FSDKRTTPEKIKDNLAGPREBELK-----AVTENTESEKQITSGSLOESKESLSNK 126  
Db 278 LQQLTAABQRNOILSQLOQMAEHNLTNTVTEREESKILLEKELEVAERKLSFH- 336  
QY 127 TVPSTSNWEICDFITKGNLVL-----SKSGVEKLSQTDHLVLPQAAQTQLIQV 178  
Db 337 NLQSEMHLLQEPQAGQAQAELESRYALSALBQKHKAEMEKTSHLSLQKTG---QELQS 393  
QY 179 ASPAFTPDKTAIAEYTSRAGENCEI-----SOLDVDGKEIINEGEVFNLSLLKVTIPT 233  
Db 394 ACDALKDQNSKLLQDKNEQAVQSAQTIOQLBDOLOQKSKEI-----SOFNLRLPQQ 445  
QY 234 GYKHIGQDAFVDNKNIAEVRNLPESLETISDYAFALHALKQIDLPDLNKAIGELAFFDNQI 293  
Db 446 -HETASQTSFPDVN-----EGTQVTEENIASLQKRVVEL-----ENEK 484  
QY 294 TGKLSLPRQLMRL-AERAPKSNHKTIEFRGNSLKVIGEAQFQNDLSQLMLPGLEKIE 352  
Db 485 GALLLSSTIEBELKAENEKLSQITLLEAQNRT---GEA---DREVSEISIVDIANKRS 537  
QY 353 SEAFITGPGDDHYNRVVLTGKSNPGLATENTYNPDKSLWQSESEIYITKWEEDF 412  
Db 538 SSA-----EESGOD-----VLENTFSQKHKL---SVLLEMEKAEQSEI 573  
QY 413 TYQKNSTVTG-----FSNKGLOKVRKNKNLE-----IPKQNGVT 446  
Db 574 AFLKLQLOQKKAEEADHEVLDQKEMQWEGEGIAPIKMKVLEDTGQDFPLMPHEESLP 633  
QY 447 ITEIGDNARNVDFQNTLRKYDLEE--VKLPSTIRKIGAFAPQSNLKSFEASDDLEE 504  
Db 634 AVE---KEQASTEHQSRSTSEISLNDAGVELKST-----KQDGDKLSLAVPDIGQC 681  
QY 505 KEGAFMNNRIETLEKDKLVITIGDAAPH-INHIIVLVPESVQIGR-----SAFRONGA 558  
Db 682 HQDELERLKSQILEL-----ELNPHKAQSEIYEKNLDEKAKESINLNLIIEBPKKAD 733  
QY 559 NN-----LIFWGSKVVTGLGEMAFSLNRLHLDL-----SEOKLQTEIPVQAQS 601

Db 734 NNSAFTALSEERDQLL-----SQVKLSMVTETRAQVKQLEMMNLAERQRRLDYESQTAH 790  
 QY 602 DNALKEVL-----LPASLKTIREEAFKXNHLKOLEVASALSHIAFNALDNDGDEQDNK- 656  
 Db 791 DNLTEQIHSLSIEAKSKDKVIEVL-QNELDDVQLQFSQSTLIRSL-----QSQLOKKE 844  
 QY 657 -----VVVTKHNSYALADGEHFIVDPOKLSSTIIVDLEKILKILIEGLDYSTLR 704  
 Db 845 SEVLEGAERVHRISSKVEELSQAQLELEITQMDQLL-----LEK-----KRDVETLQ 893  
 QY 705 QTQTQFRMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKA 764  
 Db 894 QTIEEKDQQT-----EISFSMTKVMQVQNEEFSLG-VEI-KTLKEQLNLLSRABE 943  
 QY 765 TKNGQLLERSINKAVLAYNNSAIIKKA---NVRKLEKELDLTLGLVBGKPLAQATWQGV 821  
 Db 944 AKKEQVEEDNEVSSGLKQNYDEMSPAGQISKEELQHEFDLLKKENQQRKQALINRK 1003  
 QY 822 YLLKTPLEPYIYGLNVYFDKSGKLIYALDMSDT-IGEGQKDAYGNPILANVDENEGYH 880  
 Db 1004 ELLQVRSLREELANLK---DESKK---EIIPLSETERGEVEE-----DKNKEYS 1047  
 QY 881 ALAVATLADYEGLDIKTILNLSQLTSIR-----QVPTAAVHRAGI 922  
 Db 1048 EKCVTSKQOEIEIYLKQTISEKEVELOHIRKOLEEKLAABEQFQALVKQMNOTLQDKTQ 1107  
 QY 923 FQAIONAAAEAEQLPKPTHSEKSSSES-ANSKDRGLQSNPKTNRGRH 971  
 Db 1108 IDLLQAEISENQAIIOKLITSNTDASGDGVALVRETIVVPPCTGSSEH 1157

## RESULT 16

ID ADE56033 standard; protein; 3259 AA.

AC ADE56033;

DT 29-JAN-2004 (first entry)

DE Human Protein CAAS3052, SEQ ID NO 1872.

KW Human; pain; neuronal tissue; gene therapy;

KW spinal segmental nerve injury; chronic constriction injury; CCI;

KW spared nerve injury; SNI; Chung.

OS Homo sapiens.

PN WO2003016475-A2.

PD 27-FEB-2003.

PP 14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GEO ) GEN HOSPITAL CORP.

PA (FARB ) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

DR GENBANK; CAAS3052.

XX New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

PS Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also

CC claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. Gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 3259 AA;

Query Match 4.2%; Score 214.5; DB 7; Length 3259;

Best Local Similarity 19.8%; Pred. No. 0.0016;

Matches 212; Conservative 168; Mismatches 427; Indels 263; Gaps 45;

QY 20 HNOEVSFLVKPEILKQTOASSISGADYAESGSKSLKI-----NETSGPVDVTVDL 72  
 Db 233 HEDELLQV-----TQAD-----VETEMQQLRVLRKLEHEESLVGRAQVVDL 277  
 QY 73 FSDKRTTPEKIKDNLAKGPREQELK-----AVTENTESEKQITSGSQLEQSKESLSLNK 126  
 Db 278 LQQLTAAEQRNQILSQQLQOMEAHEHTLRNTVETREESKILLEKMELEVAERKLSFH- 336  
 QY 127 TVPSTSNWEICDFTYKGNLVL-----SKSGVEKLSQTDHLVLSQAADGTQLIQV 178  
 Db 337 NLQEMHLLLEQFEQAGQAQAELESRYSALEQKHKAEMEETSHLSLQKTG---QELQS 393  
 QY 179 ASFAFTPKKTAIAEYTSRAGENGRI-----SOLDVDGKEIINEGEVENSYLLKKVTIPT 233  
 Db 394 ACDALKQNSKLLQDNEQAVQSAQTIQOQLEDQLQKSKEL-----SQFLNPLDPLQQ 445  
 QY 234 GYKHIGQDAFVDNKNIAEVLNPESLETISDYAFAHALKQIDLPDNLKAIGELAFDNDQI 293  
 Db 446 -HETASQTSFPDVYN-----EGTQAVTEENIASLQKRVVEL-----ENEK 484  
 QY 294 TGKLSLPLQMLRL-AERAFKSNHIKTIIFRGNLSKVIGEASFQDNDLSQMLPDGLEKTE 352  
 Db 485 GALLLSSTIELEBELKAENEKLSQITLLEAQNR-----GEA---DREVSEISIVDIANKRS 537  
 QY 353 SEAFTEGPDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLWQESFEIDYTKWLEDF 412  
 Db 538 SSA-----BESGD-----VLENTFSQKHKL-----SVLLLEMKEAQBEI 573  
 QY 413 TYQKNSVTG-----FSNKGLOKVKENKNLE-----IPKQHNGVT 446  
 Db 574 AFLKLOLQKRAEABADHEVLQKEMKQMEGEGIAPIKMKVFLDTGQDFPLMPNBSLIP 633  
 QY 447 ITEIGDNFRNVDFQNKTLRKVDLEE--VKLPSTIRKIGAFQSNLNKSPFASDDEE 504  
 Db 634 AVE---KEQASTEHQSRSTSEISLNDAGVELKST-----KQDGDKLSAVIDIGOC 681  
 QY 505 KEGAFMNNRIETLELKDKLVITGDAAFH-INHIYAIVLPESVQEIQR-----SAFRNGA 558  
 Db 682 HQDELERLKSQILEL-----ELNFHKAQEIYEKNLDEKAKESLNLNLEEFKKNAD 733  
 QY 559 NN-----LIFMGSKVKTGLGMAFLSNLEHLDL-----SEQKLTETIPVQAFS 601





```

Db      503 TGWVKDKGLWYYLNEGSGMATGWVKDKGL 531

RESULT 18
ABP81493
ID ABP81493 standard; protein; 621 AA.
XX
AC ABP81493;
XX
04-MAR-2003 (first entry)
XX
DT DT
DE DE
DE DE
XX Streptococcus pneumoniae polypeptide SEQ ID NO 410.
XX
KW Streptococcus pneumoniae; infection; otitis media; antibacterial;
KW diagnosis; gene therapy.
XX
OS Streptococcus pneumoniae.
XX
XX WO200283855-A2.
XX
PD PD
XX 24-OCT-2002.
XX
XX 12-APR-2002; 2002WO-US011524.
XX
PR 16-APR-2001; 2001US-0283948P.
PR 18-APR-2001; 2001US-0284443P.
XX
XX (AMCY ) AMERICAN CYANAMID CO.
XX
XX Zagursky RJ, Masi AW, Green BA, Chakravarti DN, Russell DP;
PI Wooteers JL;
PI
XX
DR WPI; 2003-093010/08.
DR N-PSDB; AB242341.
XX
XX New Streptococcus pneumoniae polynucleotides, useful for treating or
PT preventing S. pneumoniae infections, or non-systemic diseases, e.g.
PT otitis media, which are induced or exacerbated by S. pneumoniae.
XX
XX Claim 42; Page 677-680; 1091pp; English.
XX
XX The invention relates to isolated polynucleotides (AB272147-AB242522) of
CC a Streptococcus pneumoniae genomic sequence, a fragment or degenerate
CC variant of the polynucleotide or a nucleic acid sequence 95% identical to
CC one of the polynucleotides. The S. pneumoniae polynucleotides and encoded
CC polypeptides (ABP81299-ABP81674) are useful for treating or preventing S.
CC pneumoniae infections or non-systemic diseases, e.g. otitis media, which
CC are induced or exacerbated by S. pneumoniae. These are also useful for
CC detecting S. pneumoniae in a biological sample or diagnosing S.
CC pneumoniae infection in a subject. The polynucleotides have antibacterial
CC activity and are useful in gene therapy
XX
SQ Sequence 621 AA;

Query Match 4.2%; Score 213.5; DB 6; Length 621;
Best Local Similarity 23.2%; Pred. No. 0.00019;
Matches 132; Conservative 75; Mismatches 175; Indels 187; Gaps 31;

QY 6 KTVALTTLTVSVV-----THNQEVF-----SLVKEPILK-----Q 35
DB 3 KTIILSLTAAVILAAVFPNEPIADTPSEVIKETKVGSIQQNNIKYKVLTVENIGT 62
QY 36 TQASSISGADY-AESSGK-----SKLKINETSGPVDDTVTDLFS-----DK 76
DB 63 VQVNGVTPVEFEAGQGGKPTPTKITGTGDKVFTVEVASQAFSYDPETGRIVYPSS 122
QY 77 RTTPEIKDNLAKGPREQELKAVTENSEKQITSGSQLEQSKS-----120.
DB 123 ITIPSSIKKIQKGFHGSRAKTI-----FDKGSQLEKIEDRAFDFSELEEELPAS 174
QY 121 -----SLSLNK-TVPSTSNWEICDPTTKGNTLVGLSKSGVEKLSQTDHLVLP 168
DB 175 LEYIGTSAFSAFQKLLKLTFFSSSKLEL-----ISHEAFANLSNLEKLTLPKS 222

```

---

```

QY 169 AAD-CTOLIQA-----SFA-----FTPDKTAIABYTSRAGENBISQID 208
DB 223 VKTLGSLNPLRLTSLKHVDVBEKNESFASVGVLFSD-KTQLIYVPSQ--KNDESYP 279
QY 209 VDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVKNKNAIAEVLNLPESLETISDYAF-A 267
DB 280 KETKELASYSFNKNSY-LKCLELNEGLEKIGTFAFADAIKLBEISLPNSLETIERLAFYG 338
QY 268 HLAKQIDLPNLKAIGELAF-----PDNOITGKLSLPRQLMRLAERAFKSNHIK 317
DB 339 NLELKEILPDNVKNFGRHVMNGLPKLSLITGNIN---SLPSFSLGVLSLKEIHIK 395
QY 318 -----TIEFRGNS--LKVIGESFQ-DNDLSQMLPDGLEKIKISEAPT 357
DB 396 NKSTEPSVKKDTFAIPETVKFVVTSEHIKDVLSNLSNDI-----IVEKVDNIQETDV 451
QY 358 GNPDDHNNRVVLWTK-----SGKNPGLATENTYVNPDKSLW---QESPEIDY 404
DB 452 AKP-KKNSNQGVGVKDKGLWYYLNEGSGMATGWV-----KDKGLWYYLNEGSGM-A 502
QY 405 TKWLEED--FTYQKNS---VTGF-SNKGL 427
DB 503 TGWVKDKGLWYYLNEGSGMATGWVKDKGL 531

RESULT 19
ADM92258
ID ADM92258 standard; protein; 621 AA.
XX
AC ADM92258;
XX
DT 03-JUN-2004 (first entry)
XX
DE S pneumoniae antigenic protein sequence SeqID455.
XX
KW antibacterial; gene therapy; Streptococcus pneumoniae infection;
KW antigenic.
XX
OS Streptococcus pneumoniae.
XX
PN WO2004020609-A2.
XX
PD 11-MAR-2004.
XX
PF 02-SEP-2003; 2003WO-US027401.
XX
PR 30-AUG-2002; 2002US-0407082P.
XX
PA (TUFT ) UNIV TUFTS.
XX
PI Camilli A, Hava DL;
XX
DR WPI; 2004-239189/22.
DR N-PSDB; ADM92021.
XX
PT New Streptococcus pneumoniae nucleic acid molecules, useful for
PT diagnosing, treating and preventing active infections of Streptococcus
PT pneumoniae.
XX
PS Claim 27; SEQ ID NO 455; 123pp; English.
XX
CC This invention relates to novel isolated Streptococcus pneumoniae nucleic
CC acid molecules and the antigenic polypeptides encoded by them. The
CC invention may be useful for the production of compounds with an
CC antibacterial activity or for gene therapy. The nucleic acid molecules,
CC compositions and methods disclosed are useful for treating Streptococcus
CC pneumoniae infection. The present sequence is that of an S pneumoniae
CC protein of the invention.
XX
SQ Sequence 621 AA;

Query Match 4.2%; Score 213.5; DB 8; Length 621;

```



Best Local Similarity 23.2%, Pred. No. 0.00019;							
Matches 132; Conservative 75; Mismatches 175; Indels 187; Gaps 31							
QY	6	KTVALTLTTSVV-----THNQEVF-----SLVKEPILK-----Q 35					
Dd	3	KTITLSLTTAAVLAAAYPNBPILADTPSSEVIKETKVGSIIQQNNIKYKVLVEGNIGT 62					
QY	36	TQASSISGADY-AESSGK-----SKLKINETSFPVDDTVTDLPs-----DK 76					
Dd	63	VQVGNGVTPEPEAQDQKPFTPTKITVGDKVFTEVASQAFSYYPDGTGRIVVPSS 122					
QY	77	RTPPKIKDNKLAKPREOLKAVENTESEKKQTSGSQLEQSKE-----120					
Dd	123	ITIPSSIKKIQQKHFGHSKATII-----PDKGSOLEKDRAFDSELEIELPAS 174					
QY	121	-----SLSLNK-TVSTNSTNWEICDPI TKGNTLVGLSGSGVEKLSQTDHLVLP SQ 168					
Dd	175	LBVYGTSAPSPSQKLKLLTFSSSKLEL-----ISHAEPANLSNLEKTLPLKS 222					
QY	169	AAD-GTQLIQVA-----SEA-----FTPDKKTAETSYSTRAGENGESIQSD 208					
Dd	223	VKTLSNLFRLLTSLKHVDVEEGNSPASVDGVLFSD-KTOLIYYPSQ--KNDESYKTP 279					
QY	209	VDGKRIIEGEVFN SYLLKKVTITPGYKHIGQDAFVNKNIAE VNLPESLETISDYAF-A 267					
Dd	280	KETNELASYSFNFKNYS-LKKLELNEGLEGKTGFAPADAIKLEBISLFPNSETIERLAFYG 338					
QY	268	HIALKQIDLDPNLKAIGELAF-----FDNQITGKLSLRQMRLAEARFAKSNHIK 317					
Dd	339	NLELRELILLPDNVKPFGRHVNGUPLKSLTLGNVIN---SLSPFLSGVLDSLKEIHK 395					
QY	318	-----TIETPRGNS-LKVIGEASFQ-DNDLSQMLMPOGLEKIBSEAPT 357					
Dd	396	NKSTEFSVKKOTFAIPETVKFYPTVSEHKKV LKSNLSSTNDI----IVEKVONIKQETDV 451					
QY	358	GMPGDHYNRNRYVLWTK-----SGKNPSGLATENVVPDKSLW---QESPEIDI 404					
Dd	452	AXP-KGNSNQGVGHVKKDKGLWYLINSGSMATGW-----KDKGLWYLYNESGSM-A 502					
QY	405	TKWLBED----FTYQKNS-----VTGP-SNKGL 427					
Dd	503	TGWYDKDLGLWYLYNESGSMATGWYKDKGL 531					

**RESULT 20**

ADH87293	
ID	ADH87293 standard; protein; 1799 AA.
XX	
AC	ADH87293;
XX	
DT	22-APR-2004 (first entry)
XX	
DE	Enterococcus faecalis polypeptide #1773.
XX	
KW	Enterococcus faecalis infection; transcription regulatory element; antibacterial.
XX	
OS	Enterococcus faecalis.
XX	
PN	US6617156-B1,
XX	
PD	09-SEP-2003.
XX	
PF	13-AUG-1998; 98US-00134000.
XX	
PR	15-AUG-1997; 97US-0055778P.
XX	
PA	(DOUC/) DOUCETTE-STAMM L A.
PA	(BUSH/) BUSH D.
XX	
PI	Doucette-Stamm LA, Bush D;
XX	
DR	WPI; 2003-895394/82.

DR N-PSDB; ADH83888.  
XX  
PT New nucleic acid comprising a sequence encoding an *Enterococcus faecalis*  
PT polypeptide, useful for preparing a composition for diagnosing or  
PT treating *E. faecalis* infection.  
XX  
PS Disclosure; SEQ ID NO 5178; 193pp; English.  
XX  
CC The invention relates to *Enterococcus faecalis* polynucleotides and  
CC polypeptides. The invention also relates to a recombinant expression  
CC vector comprising a polynucleotide operably linked to a transcription  
CC regulatory element, a cell comprising a recombinant vector, a method for  
CC producing an *E. faecalis* polypeptide, an isolated nucleic acid comprising  
CC a sequence not given in the specification, a recombinant vector  
CC comprising the nucleic acid and a cell comprising the recombinant vector.  
CC The polynucleotides can be used to detect the presence of *E. faecalis* in  
CC a sample. The sequences are useful for preparing a composition for  
CC diagnosing or treating *Enterococcus faecalis* infection. This sequence  
CC represents an *E. faecalis* polypeptide of the invention.  
CC

[illegible]

QY 555 QN:---GANNLIIFMGSKVKTGLGEMAFSLNRLHLDLSEKQOLTEIPV---QAFSDNALK-- 606  
D 1335 QTVDSAAAELEKASQLEKMPDVA---NKAD-LEKAIQEGLAGKPKSDQEBTEETKKVL 1390  
QY 607 EVLLPASPALKTIREEAFKQKHLLKQLEVASALSHIAFNALDDNDGDBQFDKNNVVKVTHNSY 666  
D 1391 EESLAAQKVFQAEKVTOBEIDQ-----ATKTLREAI 1422  
QY 667 ALADGEHFTVDPKLSSSTIVDLKILKLTLEGLDYSLTROTTOPTQPRDMTACKALLSKSN 726  
D 1423 AQLKEQPVAVDKETLKEQIAQ-ARGKPEEGYQFT---KETEKQLEAIQAAEAIVAKET 1478  
QY 727 LRQGEKQKFLQBAQFPLGRVLDKAIKAEKALVTKKA-----TKNQQLLERS---INKA 778  
D 1479 ATKEEVSEALNALETAMAQLKEVPLVNDKQLQEVVKRAQQVTPSEGHQFTASSLQELQKA 1538  
QY 779 VLAYNNSAIKK--ANVKRLEKELDLTLGLVEGKGLAQATWVGVLTKTPPLPEYY-- 834  
D 1539 LLAARKNT-LKNPAANQKMIDEAAVELTSAIDG---LQEBVLVTDKKALEAMIAKAKAIKP 1594  
QY 835 -LGLNVYFDKSKLIVALDMSDTIGEGQKDAYGNPILNVDENEGVHALAVATLADYEG 893  
D 1595 SAGKEFTSESKARLTAIDQAEGL-LADKNARQEQI-DIAEKN-----VKTALD--SL 1643  
QY 894 DIKTILNSKLSQTSIRQVPTAAVHRAG--IFQAIQNAAE---AEQLLPKPGTHSEKS 947  
D 1644 E-BQVLQTDKTKLKEQLQAEETLKPKAGKQFTKASQEAALAEAIKQAKALVEDPNATQEA 1702  
QY 948 SSESANSK-----DRGLQSNKPNRGRHSAI----- 974  
D 1703 DKCLSLQSAIEAMAEPISSNSTGNNGHSTVSGTSGKGTATGTTTKTTTSGT 1762  
QY 975 LPTGSKGSFVYGILGY---TSVALLSLITAKKK 1007  
D 1763 LPKANVSVSPISGFLIVSGLKLPFFKNKQ 1798

## RESULT 21

ABU46539 standard; protein; 2045 AA.  
XX AC ABU46539;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #32066.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Streptococcus pyogenes.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX DR WPI; 2003-029926/02.  
XX DR N-PSDB; ACA50409.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
PS Claim 25; SEQ ID NO 74463; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

Sequence 2045 AA;

Query Match 4.2%; Score 213; DB 6; Length 2045;  
Best Local Similarity 18.9%; Pred. No. 0.0011;  
Matches 245; Conservative 193; Mismatches 466; Indels 392; Gaps 59;

QY 29 KEPILKQTOQASSISGADYAEBS-GSKXLKNETSGPVDVTDVDFSKRTTPEIKDNL 87  
D 822 KERLKSDTQKVDKAKNDAIKAEPEGKVNIPQAHIP-----GDLNKK-----EKLAE 872  
QY 88 AKGPREQL-----KAVTENTESEKQITSGSQLEQSKESL-----SLNKTVPSTSNWE 135  
D 873 KQKADDTTEKAIQVDTLTKLDEKQKQVTKAELEKAKTDVKNQTPRELDKKVPELKK-A 931  
QY 136 ICDFTIKGNTLVGLSKSGVEKLSQT-----DHLVLPQADGTQLQVASF 181  
D 932 IEDTHVKN-LEGVKNKAIEDLKKHAHTETAKINGDDTLDKATKEADKQKALAAAGKD 990  
QY 182 AFT----PDK-KTAIAEYTSR---AGENGELISOLDVDCGEIINE-GEVFNLSLLKKVTIP 232  
D 991 AITKADDADKRVSTAVTEHTPKIAAHTGDDKKAQVQDANTALDKAAEKEGEINKDATUT 1050  
QY 233 T--GYKHIGQ-----DAFVD--NKNIAEYN----- 253  
D 1051 TEDKAKQLKEVETALTAKQNVKAAKTADAINDARDKGVATIDAVHKAGQDLGARKSGQV 1110  
QY 254 --LPESLETISDYAPAHUALKQIDLPDLNKA-----GE-L 286  
D 1111 AKLBEAAKATKDKISADPTLTSKEKEQKQVDAELKKAIEAVNAADTADKVDDALGEGV 1170  
QY 287 AFFDQITGKLSLP-----ROLWRLAER---AFKSNHIKTIEPRGNSLVKIGEA---- 332  
D 1171 TDIKNQHKSGSDIDARRREAHGKGLDRVAQETKGAIEKDPFTLTTEBAKQKVDVDAKBERG 1230  
QY 333 -----SFDQNDLSQLMLPDGLEKIESEAFNGPDGDH---YNNRVVLWTKSGKPNGLAT 384

Db 1231 MAKLEAKDADALDKAYEGVGTDIKNQHKSGDPDVARGLHNKSIDEVAQATKD--AITA 1288  
QY 385 ENTYNPKSLWQSPEDYTKWLEB-----DFTYQKNSVTGNSKGLQKVKNK 434  
Db 1289 DTTLTEAEKQRTQGNVDKEATKAKELAKADADALDKAY-GDGVTSIKNQH-----KSGK 1343  
QY 435 NLEIPKQHGVTITIEIGNAFNVDQ--NKTLRKYDLEEVGLPSTIRKIGAFAPQSN 491  
Db 1344 GLDVRKDEHKALEAVAKVTAIEADPTLPEVREQQAEVQ----- 1386  
QY 492 LKSPF-ASDDLEBEIKEGAFMNNRIETLEKOKLVLTIGDAAFHNIHAIYILPESVQIEGR 550  
Db 1387 -KELELATDKIAEAKDA-----DEADKAYGDGVTAIENAHVIGKI-EARKOLAK 1434  
QY 551 SAFRQNGANNLIFM-----GSKVTILGEMAFLSNRLEHLDLSEOKQLTEPIPVQFAS 601  
Db 1435 KDLAEAAAKTKALITIEDKTLTDQKEQLLGVDTTEYAKGIENIDAADAAAGVD---KAYS 1491  
QY 602 D-----NALKEVLLPAS--LKTIREEAF-----KKHHLKOLEVAS 634  
Db 1492 DGVRIILAQYKEGQNLNDRNNAKEFLLEKADKVTKLINDPTLTHDQVQDQINKVEQAK 1551  
QY 635 ALSHIAFNALDNDGDEQFDNKV--VVKTHNSYALADGEHFIVDPDKLSSTIVDLKIL 692  
Db 1552 L---DAIKSVDDAQTADALDALKGKIENINNOYOHGDG---VDVRKATAK-GDLEKEA 1603  
QY 693 KLIIEGLDSTLRQTQTQF-RDMTTA-----GKALLS-K 724  
Db 1604 AKVKAL---IAKDPTLTQADKQTAQAAVDAAKNTAIAAVDKATTTEGINELGKGITAIN 1660  
QY 725 SNLRGEXKQKPLQEAQFFLGRVLDLAKAKAEKALVTK-----KATKNGQL----- 770  
Db 1661 KAYRPEGVGVKARKEA-----AKADLEKEAKV-KALITINDPTLTAKADKAKQTEAVAKALKA 1715  
QY 771 -----LERSINKAVLAYNNS-----AIKKANVKELEKELDILLTGLVEGK 809  
Db 1716 AIAAVDKATTABGINQELGKGITAINKAYRPEGVGVKARKEAKADLEREAQVREAIAND 1775  
QY 810 GPLAQTWVQGVYLLKTPILPLPEYI-----GLNVYFDKSGKLIYALDMSDITGEG-- 860  
Db 1776 PTLTKRADKAKQTEAVAKALKAIAAIVADKATTAEAGIN--QELGKGITAINKAYRPEGVE 1832  
QY 861 -----OKDAVGNPILNVDENEGHALAVAT 886  
Db 1833 AHKEAAKANLEKAKETKALISGRYLDSETEKAVQKQAEVQALAKALQVEAAKTVEAVK 1892  
QY 887 LADYEGLDIKTILNKSOLTSIRQVPTAAVYHRA--GIFQATQNAABE--AEQLLPKPGT 942  
Db 1893 LA--ENLGTVAIRSVAAGLAKDTQATPAALNEAKQAIEALKQAAAEATLAKITTDAKLT 1950  
QY 943 HSEKSSSSSESAN-----SKDRGL-----QSNPKTNRRGH 971  
Db 1951 EAQKAEQSENVSLALKTAIATVRSQAISVKEAKDKGITAIRAAVYVFNKAVAKSSSANH 2010  
QY 972 SAILPRTSGSFVYGIIGYTSVALLSLITATKKK 1007  
Db 2011 ---LPKSGDANSIVLGLGWSL-LLGWVLYSKKKE 2042

RESULT 22  
ADR83918  
ID ADR83918 standard; protein; 2045 AA.  
XX  
AC ADR83918;  
XX  
DT 02-DEC-2004 (fiset entry)  
XX  
DE S. pyogenes hyperimmune system reactive antigen Spy0737.  
XX  
KW hyperimmune serum reactive antigen; vaccine; anticaline.  
XX  
OS Streptococcus pyogenes.  
XX

PN WO2004078907-A2.  
XX  
PD 16-SEP-2004.  
XX  
PF 02-MAR-2004; 2004WO-EP002087.  
XX  
PR 04-MAR-2003; 2003EP-00450061.  
XX  
PA (INTE-) INTERCELL AG.  
XX  
PI Meinke A, Nagy E, Winkler B, Gelbmann D;  
XX  
DR WPI; 2004-653698/63.  
XX  
DR N-PSDB; ADR83768.  
XX  
PT New isolated nucleic acid molecules encoding hyperimmune serum-reactive  
PT antigens from Streptococcus pyogenes, useful for diagnosing, preventing  
PT and treating S. pyogenes infections.  
XX  
PS Claim 13; SEQ ID NO 186; 145pp; English.  
XX  
CC This invention describes a novel nucleic acid molecule encoding a  
CC hyperimmune serum reactive antigen or its fragment from Streptococcus  
CC pyogenes. The nucleic acid molecule or hyperimmune serum-reactive antigen  
CC or its fragment are useful for the manufacture of a pharmaceutical  
CC preparation, especially a vaccine, against S. pyogenes infection. In  
CC addition, the hyperimmune serum reactive antigen or fragment is used for  
CC the isolation and/or purification and/or identification of an interaction  
CC partner of the hyperimmune serum reactive antigen or its fragment, for  
CC the generation of a peptide (e.g. anticalines) binding to the antigen or  
CC fragment, or for the manufacture of a functional nucleic acid selected  
CC from aptamers and spiegelmers. The nucleic acid molecule may also be used  
CC for the manufacture of functional ribonucleic acids, such as ribozymes,  
CC antisense nucleic acids and siRNA. ADR83733-ADR84189 represent S.  
CC pyogenes hyperimmune serum reactive antigens, fragments and the encoding  
CC polynucleotide described in the invention.  
XX  
SQ Sequence 2045 AA;

Query Match 4.2%; Score 213; DB 8; Length 2045;  
Best Local Similarity 18.9%; Pred. No. 0.0011;  
Matches 245; Conservative 193; Mismatches 466; Indels 392; Gaps 59;  
QY 29 KEPILKQTQSSSISGADYAESS-GSKLKLNETSGPDDVTVDLFDKRTTPPKIDNL 87  
Db 822 KERLKSOTQKVKDAKNADATKKAPEEGKVNIPQAHIP-----GDLNKDK-----EKLAE 872  
QY 88 AKGPREOEL-----KAVNTETSEKQITSGSQSKESL-----SLNKTVPSTSNWE 135  
Db 873 KQKADDTTEKALDVDTLTDEBEKQKVKTKAELEKAKTDVKNQTTRELDKKVPELKK-A 931  
QY 136 ICDPITKNTLVGSGSGVEKLSQT-----DHLVLPQAADGTQLIQVASF 181  
Db 932 IEDTHVXGN-LEGVKNKAIEDLKKAHTEVAKINGDDTLDKATKEAQVKEADKALAAAGD 990  
QY 182 AFT-----PDK-KTAAIEYTSR---AGENGESQLDDVGKEIINE-GEVFNYSLLKKVTIP 232  
Db 991 AITKADDADKSVTAVTEHTPKIKAAHKTGDLUKQAQVDANTALDKAAEKEGEINKDATLT 1050  
QY 233 T--GVKHIGO-----DAFVD--NKNIAEVN----- 253  
Db 1051 TEDRAKQKLVETALTAKONVKAAKTADALINDARDKGVATIDAVHKAGDGLGARKSQV 1110  
QY 254 --LPESLETISDYAFAPHALAKQIDLPDNLKAI-----GE-L 286  
Db 1111 AKLEEAATKTKDISADPTLTSTKEKEQSKAVDAELKKAIEAVNAADTADKVDDALGEGV 1170  
QY 287 AFFNQITGKLSLP-----RQLMRLAER---AFKSNHIKTIIEFRGNSLKVIGEAE---- 332  
Db 1171 TDIKNQHKSGDSIDARREAHGKELDRVAQETKGAIEKDPDTLTTEBAKQKVDVDAAKEERG 1230  
QY 333 -----SFQDNDSLQMLPDGLEKIESAFTGNPGDDH---YNNRVVLWTKSGKPNPSGLAT 384



Db 1125 AKLEAAKATKDKISADPTLTSTKESEKSKAVDAELKKAIEAVNAADTADKVDDALGEGV 1184  
QY 287 AFFDNQITGKLSLP-----RQLMRLAER---AFKSNHIKTIYPRGNSLKVIGBA---- 332  
Db 1185 TDIKNQHKSGSIDARREAHGKELDRVAQETKGAIEKDPPTLTTEKAKQKVDVDAKERG 1244  
QY 333 -----SPQNDLSQLMLPDGLEKISEAFTGNPGDDH---YNNRVVLMTKSGKNPSGLAT 384  
Db 1245 MAKLEAKADADALDKAYGEGVTDIKNQHKSGDFVDARRGLHNKSIDVAQAATKD--AITA 1302  
QY 385 ENTYPNPKSLWQESPEIDYTKWLEE-----DFTYQKNSVTGFSNKGLOKQKKNK 434  
Db 1303 DTTLTEAEKETQGNVDEKATKAKELAKADADALDKAY-CDGVTSTKNQH-----KSGK 1357  
QY 435 NLEIPKQNGVVTITIGONAFRNVDFQ---NKTLYDLERVLKPLSTIRKIGAFAFQSN 491  
Db 1358 GLDVRKDEHKKALEAVAKRVTAIEADPTLTPEVREQQAEVQ----- 1400  
QY 492 LKSPF-ASDLEIEKEGAFMNNRIETLKKLVITIGDAAPHINHIYAIVLPESVQETGR 550  
Db 1401 -KELELATDKTAEAKDA-----DEADKAYGDVTAIENAHVIGKGI-EARKOLAK 1448  
QY 551 SAFRQNGANNLIFM-----GSKVKTIGEMAFLSNRLEHLDLSEQKQLTETIPVQAFS 601  
Db 1449 KDLEAAAKTKALIIEDKTLTDQKQQLGVDTYVAKGIENIDAKDAAGVD---KAYS 1505  
QY 602 D-----NALKVLLPAS---LKTIRSEAF-----KXNHLKOLEVAS 634  
Db 1506 DGVDRILAQYKQGNLDRNNAKEFLKKEADKVTKLINDPTLTHDQKVDQINKVEQAK 1565  
QY 635 ALSHTAFNALDDNDGQFNDKV--VVKTHNSYALADGEHFVDPDKLSSTIVDLKIL 692  
Db 1566 L--DAIKSVDDAQAADALGKGIENINNOYQHGDS---VDVRKATAK-GDLEKEA 1617  
QY 693 KLIEGDYSLRTQTQTFQ-RDMTTA-----GKALLS-K 724  
Db 1618 AKVKAL---IAKDPTLTQADKQKTAANDAAKNTAIAVDKATTTGEGINQELGKITAIN 1674  
QY 725 SNLRGKQKQFLOEAFQFGRVLDLKAIAKAEKALVTK-----KATKNGQL----- 770  
Db 1675 KAYRPGEGVKARKEA---AKADLEKEAAKV-KALITNDPTITKADKAKQTEAVAKALKA 1729  
QY 771 -----LERSINKAVLANN-----AIKKNVKELEKELDLTLGLVEGK 809  
Db 1730 AIAAVDKATTAGINQELGKITAINKAYRPGEGVKARKEAAKADLEKEAAKVRREAIAND 1789  
QY 810 GLAQATWQGVYLLKTLPLPEYI-----GLNVYFDKSGKLIYALDMSDTTIGEG-- 860  
Db 1790 PTLTKADKAKQTEAVAKALKAIAAIVDKATTAGIN---QELGKITAINKAYRPGEGVE 1846  
QY 861 -----QXDYGNPILNVDEDNEGYHALAVAT 886  
Db 1847 AHKKAANLEKAVAKETKALISGDRYLSGTEKAVQKQAEQALAKALQVEAAKTVEAVK 1906  
QY 887 LADYEGLDIKTILNSKLQSLTISIRQVPTAAHYRA--GIFQATONAAE--AEQLLPKPGT 942  
Db 1907 LA--ENLGTVAIRSAYVAGLAKDTDOATAALNEAKQAAIEALQKAAETLAKITTTDAKL 1964  
QY 943 HSEKSSSESAN-----SKDRGL-----QSNPKTNRGRH 971  
Db 1965 EAQKAEQSENVSLAKTALATVRSQAOSTASVKEAKDKGITARAAYVFNKAVAKSSANH 2024  
QY 972 SAILPRTSGKGFVGIIGYTSVALLSLITAKKKK 1007  
Db 2025 ---LPKSGDANSIVLGLGVMSL-LLGWLVYSKKKE 2056

## RESULT 24

ABP73809

ID ABP73809 standard; protein; 1881 AA.

XX

AC ABP73809;

XX

DT 30-JAN-2003 (first entry)  
XX Candida albicans essential protein SEQ ID NO 7646.  
DE Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;  
KW signal transduction; DNA replication; cell division; growth;  
KW proliferation; Candida albicans; fungicide; antifungal.  
XX  
OS Candida albicans.  
XX WO200253728-A2.  
PN 11-JUL-2002.  
PD 26-DEC-2001; 2001WO-US049486.  
PF 29-DEC-2000; 2000US-0259128P.  
PR 20-FEB-2001; 2001US-00792024.  
PR 22-AUG-2001; 2001US-0314050P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
PI WPI, 2002-566694/60.  
DR N-PSDB; ABZ33359.  
XX  
PT Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele of  
PT a gene and placing other allele of the gene under conditional expression.  
XX  
PS Claim 44; SEQ ID NO 7646; 167pp + Sequence Listing; English.  
XX  
CC The invention relates to constructing (M1) a strain of diploid fungal  
CC cells in which both alleles of a gene are modified, comprising modifying  
CC one allele by insertion or replacement by a cassette having an  
CC expressible selectable marker and modifying other allele by  
CC recombination, of a promoter replacement fragment with a heterologous  
CC promoter, so that expression of the second allele is regulated by the  
CC promoter. (M1) is useful for constructing a strain of diploid fungal  
CC cells in which both alleles of a gene are modified. The diploid fungal  
CC cells having both alleles modified are useful for identifying a gene that  
CC is essential to the survival or growth of a fungus, a gene that  
CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
CC that contributes to the resistance of a diploid fungus to an antifungal  
CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
CC and for identifying a therapeutic agent for treatment of a mammalian  
CC disease. (M1) is useful for identifying a compound which modulates the  
CC activity of a gene product, preferably enzymatic activity, carbon  
CC compound catabolism, biosynthetic, transporter, transcriptional,  
CC translational, signal transduction, DNA replication and cell division  
CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office  
SQ Sequence 1881 AA;

Query Match 4.2%; Score 211; DB 5; Length 1881;  
Best Local Similarity 19.9%; Pred. No. 0.0012;  
Matches 227; Conservative 173; Mismatches 431; Indels 310; Gaps 48;

QY 5 LKTVALTUTTVSVVTHN-----QEVFSLVKEPILKQTOASSSIS 43  
Db 689 LPTVLTVDVYVNLVKDNFLRIKRALFHPDPSPRGKISYELFEEDTKVNLTKELQ--T 746  
QY 44 GADYAESGSKGIKINETSQGVDDTVTDLFSQDKRTTPPEKIKDNLAGPREQELKAVTNT 103  
Db 747 EKENAESNDK---ELNEKI-----EKLTNLSTKLET---KLED-----KEQELAKIQSDH 790





```
Db 403 YDAEELNNDKSTAKVLIEINLSEVKHNLSEITNIKQGGKIKYKAKIMQIKATSE 462
QY 359 NPG-----DDHYN-----NRVLWTGKGNPGLATENTVVPNDKSLMOES 399
Db 463 NPAKTELEKVDQSNVYNYLNIQTERRNLIVTEKNRLN-----GIDSTITIEGAL-KES 517
QY 400 P---BIDYTKWLEE-----DPTYQK-NSVTG-----PSNKGLOKVKRKNL----- 436
Db 518 KGNYSIGLEKLEEGIKRKLKVDITKKSINSVTGNFSSLFNFDLNIQYDFNKNINDYEN 577
QY 437 ---EIPKHGVTITEIGDAPFRNV-----DFQN-KTLR-----KYDL-----EEVKLP 476
Db 578 KNGEYNEPEG-SLNKISEN-LRNASENTSDYNSAKTLLEBAQKEKVNLLNKEEANKYL 635
QY 477 STIRKIGAFAPSNLKLSPFASDDLEETKEGAFMNRRIETLELKDVLVTI-----GDAAPH 532
Db 636 RDVKVSGFRP-----IFNMKESLDKINE-----MIKKEQLTVNEGHGNVKQL 678
QY 533 INHIYAVLPESVQIGRSAPFONGANNLI-----FMGSKVKTILGEMAFPSNRLEHLDLSEQ 589
Db 679 VENIKELVDENNLSDILQATOKNEEIOKITHSTLKNKAKTI-----LGHVDTSAK 729
QY 590 KQLTIPQAFSDNALKEVLLPASLKTIREEAPK-KNHL-----KOLEVASALS 637
Db 730 -----YVGIKITPELALTELLGDAKLKTAGELKFEKNVNVLETENMKNNTNELDVKHIQ 785
QY 638 ---HIAFNALDND-----GDB-----QFDNKV-VVKTHHNSYALA 669
Db 786 DAYVALEILAHSDIEDITKQDSSKLIEMGNQIYLVKVLINGYKNKISSIKSEAVSVK 845
QY 670 DG-----EHP-----IVDPKLSSTIVDLKILKLGLOYSTLRQTT-----OTQ 710
Db 846 IGVNSKKHSELSKITCSDSKSYDNIILAEKQTEL-QNLANSFTQETNTNSDSKLEKIKTD 904
QY 711 PRDMTITAGKALLSKSN-----LROGEKQKFLQEAQFGLGRVDL 748
Db 905 FESLKNALKTLEGEVNALKASDNHEHVQSKSEPNPALSEIEK-----EETDIDSLNTAL 960
QY 749 DKAIAKAE-----KALVTKATKNGQL-----LERSINKAVLAYNNSAIKKANVKRL 795
Db 961 DELLKKGRTECVSRVKLIKDTVTKEISDDTELTINTIEKNV-KAYLAY-----IKKNYEDTV 1015
QY 796 EKELDLILGLVEGKFLAQATWQGVYLLKTPLPPEYIYGLNV-----YFDKSGK--- 846
Db 1016 Q---DVLT-----LNEHFNKQVSNHBEFTNPDKSNKSSB 1046
QY 847 -LIYALDMSDTIGEGQDAYGNPILNVDBEDNEGYPHALAVATLADYEGLDIKTILNSKLSQ 905
Db 1047 ELTKAVTDSKTIISKLKGV-----IIEVNENTE-----MNTIESSAKEIEALYNELKNK 1095
QY 906 LRSIRQV 912
Db 1096 KTSLENEI 1102

RESULT 26
AAW24575
ID AAW24575 standard; protein; 1254 AA.
XX
AC AAW24575;
XX
DT 25-MAR-2003 (revised)
DT 10-NOV-1997 (first entry)
XX
DE Merozoite apical-end protein clone 5.3.
XX
KW Merozoite apical-end protein; MAEP; Plasmodium vivax; antigen; malaria;
KW human; erythrocyte; antimalarial vaccine; antibody; blood stage parasite;
KW Duffy blood group antigen; red blood cell; therapy.
XX
OS Plasmodium vivax.
XX
```

```
PN US5646247-A.
XX
PD 08-JUL-1997.
XX
PF 04-OCT-1991; 91US-00792865.
XX
PR 05-APR-1989; 89US-00334041.
PR 06-APR-1989; 89US-00334270.
PR 03-APR-1990; 90MO-US001849.
PR 02-NOV-1990; 90US-00608639.
XX
PA (UTNY ) UNIV NEW YORK STATE.
XX
PI Galinski MR, Barnwell JW;
XX
DR WPI; 1997-362995/33.
XX
DR N-PSDB; AAT80072.
XX
PT Plasmodium merozoite apical end protein - useful as antigen for
XX production of anti-malarial vaccines.
XX
PS Claim 1; Col 29-38; 68pp; English.
XX
CC AAW24575 and AAW24576 represent the merozoite apical end proteins (MAEP)
XX isolated from two different Plasmodium vivax strains. These proteins are
XX the antigens of the invention, and immunoreact with antibodies against a
XX native MAEP sequence. P. vivax is one of the four malarial species that
XX infects humans, and is difficult to target for a vaccine, as it cannot be
XX cultured in vitro. The preinvasion orientation of malarial merozoites
XX indicates that the apical end plays an important role in the invasion
XX process. The MAEP protein binds to the surface of susceptible
XX erythrocytes from P. vivax susceptible humans and primates, and also
XX binds to rabbit erythrocytes. The antigen can be used for the production
XX of antimalarial vaccines. The antigens are involved in the invasion
XX process, and are immunochemically reactive with antibodies raised against
XX malaria (particularly P. vivax) blood stage parasites. Synthetic
XX proteins, polypeptides, peptide fragments and analogues of these antigens
XX can be used similarly. As the antigens specifically bind to a Duffy blood
XX group antigen (the antigen present on the surface of susceptible
XX mammalian red blood cells), and are necessary in the process of invasion
XX of red blood cells by merozoites, they can be used to inhibit the
XX invasion of red blood cells by a malarial organism. The antigens can also
XX be used in a method for inhibiting invasion of susceptible mammalian
XX blood cells by malarial merozoites, and in a method for inhibiting the
XX propagation of a malarial organism in susceptible red blood cells.
XX (Updated on 25-MAR-2003 to correct pp field.)
XX
SQ Sequence 1254 AA;
XX
Query Match 4.1%; Score 209.5; DB 2; Length 1254;
Best Local Similarity 21.3%; Pred. No. 0.00086;
Matches 244; Conservative 176; Mismatches 350; Indels 377; Gaps 66;
QY 23 EVFSLVKEPILKQTAASSISGAD-----YAE-----SSGSKLKINETSGPVDDTVDLFS 74
Db 76 EKILYKKEIDEIKQKTNEYKQGTSNFYTYEQYNSATQSKAKIEQ-----FINIAT 127
QY 75 DKRTTPEKIDNLAGPREQELKAVTENTSEKQIT---SGSOLEQSKESLSLNTKVPST 131
Db 128 TKKGTSDTSQD-----INELESIKBEVHKNLQLVKQESNMBEEMRKQILSKMDLLTLN 180
QY 132 SNWEICDFITKG-NTLYGLSKSGVEKLSQTDHLVLPQADGDTQLIQVASFATPDKKTA 190
Db 181 NSETIAKEISNNTQNALGFRENAKTKLTKTDEL-LQRVAA-----MIEAK-----AHQNN 230
QY 191 IARYTSRAGENGESQLDDVGKEIINEGEVFNYSLLK-----KVTIPTGYKHIGQD--A 242
Db 231 IDIALEDAQIDTEVSKIEQINREIMNKKDEIKSYLSEIKYKDKCTTEISNKGKDKIE 290
QY 243 FVD-----NK-NIAEVNLPESLETISDYAFAPHALKQI-DLPDLNKAIGEL----- 286
Db 291 FLEKFKPNEESNKNVINEIN--ENIRNSQY-----LKDIEDAKQASTKVELPFHKH 342
```





```
CC invention.
XX SQ Sequence 1905 AA;

Query Match
Best Local Similarity 4.1%; Score 207; DB 8; Length 1905;
Matches 208; Conservative 150; Mismatches 366; Indels 290; Gaps 44;

QY 29 KEPILKQTOASSISGADYAEBSG-----KSKLKINETSPPVDDTVTLDFSDKRTTP 80
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
730 KNDMTWQLESQSQAGDDHSKTRSLSEAHLKHELELNLS-----LNDLHVESKTAA 783
QY 81 EKIKONLAK-GPREQLKAVTEN-----TESKQITSGSQ-----LEQ-----SKESLUN 125
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
784 ESALQRIAELETQVQELSGAAEOSLSHLTEFESKLSAAEKSMQLEQELKDATECNSR 843
QY 126 KVPSTS-NWEICDPTTKNTL-----VGLSKSG-----VEKLSQ 159
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
844 LRVDELSELEBESSLTKGALEHATSKKIDLEALYQSILLEDETMKLOQAGENLTQKETE 903
QY 160 THLVLPQAGDGTOLIQVAFPTPKKTAIAEYTSRAGENGESQLDVGKEIINGE 219
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
904 COELSEKLKAEE---QAASY---QAKATAAE-----EVESVKVELEAFETEIS 947
QY 220 VNSYLLKKVVTPTGYKHIGQDAFYDNKNIAEVNLPESLETISDYAFALHAKQIDLPDN 279
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
948 TLETTIEELTKASNAESRAEQALVESAMMSETN-----QALKE-DLDAK 991
QY 280 LKAIQEL-AFFDNQITGKLSPLQMLRAERAFK--SNHIKTIEF-----RG----- 323
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
992 LAMLELQEQFDSHAEK-----EYFTKLSAHEKTIEHLTEVSHRGLHLATA 1040
QY 324 -----NSLKVIGEASFQNDLSQLMLPDGLE-KIHESAFTGNPGDDHNNRVV 370
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1041 ESKNAELEAQMHEALETIGKDSVKDLNERLA--ALESEIESLTHVNEAMKQEINAKVL 1098
QY 371 LMTSGKQPSGLATENTYVNPDKSLWQSPSPE---IDYTKWLEEDFTYQKNSVTFGSNKL 427
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1099 KYDELQEKLSSTSEKEAEKVVVHVEKTHIEHLREHSEGRLE-----L 1141
QY 428 QVKRNKNLEIPKQHNGVIT-----EIGDNAPRNVDFQNTLRKYDLEEVKLPSTIRK 481
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1142 QSAEERSAEIENELREVLETVAQEAETDLKEKLVLETENEKLVINEAKELGELDTK 1201
QY 482 IGAFAPQNNLKSFEASDDLEIKGAFMNNRIETLEKDKLVTTIGDAAFHNHIAVL 541
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1202 VAMF-----DELQE---QFSSTHAEKEBAEAKLAVHERTISHLTEVHTRSL 1244
QY 542 PESVQEIGSAFRQNGANNLIWFGSKVTLGEMAFLSNLEHLDLSEKQKLTETIPQAFS 601
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1245 -----ELHSAEBSKNEE-----IESKLHEALMA-AQKEAEVKLSKLDALIEELGYBE 1293
QY 602 DNALKEVLLPASLKTITREAPFKNNHLKOLEVASALSHIAFNALDNDGDGEQDNKV-VVK 660
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1294 EQATEAABAEETHKIKFDDAVIK--IKSLEEQLAVI-----ENKVELFH 1335
QY 661 THNYSALAD---GEHFIVDDPKLSSTIVDL-----EKILKLEGLDYSTLR-Q 705
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1336 TEKENLVIANSKLNELHLQNKLELQVALAAVAEKEGSEETHSRLKTLDMGIMQKE 1395
QY 706 TTQTFPRMTTAGKALLSKSNURQEKQFLEQAQFFLGRVDLDKAIKAEKALYTKAT 765
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1396 ELESQVSTVEEHEELKSKYNTLLEKQ--LLNEKYESAKEGELGAIKLEEQMNVDKSE 1453
QY 766 KNGQL--LERSNKAVLANNNAIKKANKVLEKEDLLTLGLVEGKPLAQATWVQGVYL 823
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1454 KELHLSKLERQITLSELKTMEE-IQTMQVETTEKEALTTMQE-----HANLVH---- 1502
QY 824 LKTPLEPPEYIIGLVYFDKSGKLIYALDMSDTIGSGQKDAYGNPILNVDEE-NRGYHAL 882
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1503 -----EKDALEQQLLEVRKELNDAYH-- 1523
QY 883 AVATLADYEGLDIKTILNSKLSQLTSIRQVPTAAVHR--AGIFQAIQNAABAE 934
```

Db 1524 ---TIANQE-----EQSVREIKWDAYKKFSESDQLEAEQHVAELE 1561

RESULT 28  
ABU42253  
ID ABU42253 standard; protein; 3533 AA.  
AC ABU42253;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #27780.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Staphylococcus aureus.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA46123.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 70177; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 3533 AA;

Query Match	4.1%	Score 207	DB 6	Length 3533
Best Local Similarity	19.2%	Pred. No. 0.0053		
Matches 231	Conservative 169	Mismatches 424	Indels 380	Gaps 52
Qy	18	VTHNQEVFSLVKBPILKQTKQASSISGADVAESSGSKL	-----KINETSG	-----63
Db	1515	VQAESIIAGLNPTINK	---GNVSATQAVISSKNALDOVERLAQDKQTAGNSLNHL	1570
Qy	64	---PVDDTVTDLFSDKRTTPEKIKDNLAQPREOE-LKAVTENTESEKQITSGSL	-----115	
Db	1571	QLTFAQQOALENQINNATTRDKVAEIIIAQAQALNEAMKALKESIKDPQTEASSKFINE	1630	
Qy	116	EQSKESLS	-----LNKTVPSTSNWEICDFTKGTNTLVGLSKSVEKLSQTDHLVLP	166
Db	1631	QAKQDAYTQAVQAHAKDLINKTITDPLAKSIIDQATQAVTAKNNLHGDKLQAK	-----D 1684	
Qy	167	SQAADGTLQIQVASFAPTPDKKTAIAEYTSRAGENGESIQLDVDGKEIINEGEVNSYLL	226	
Db	1685	KQRA--TETLNLNSLNTPQOQ-ALENQINNAATRGEVAQKLTQAQALNQAMEALRNSIQ	1741	
Qy	227	KKVTIPTGYKHIGOD	-----AFVNKN--IAEVNLP	269
Db	1742	DOOQTEAGSKFINKPKQDAYQAQVAKDNLINQTNPTLDKAQVEQLTQAVNQAKDNL	1801	
Qy	270	ALQKIDLPDLNKALIGELAFPDNOITGKLSLPRQLMRLAERAFKSNHIKTIFFGNSLAKVI	329	
Db	1802	HGDQKLADDKQHAVTDL	---NQLNGLNPNORQAL-----ESQINNAATRGEVAQKL	1849
Qy	330	GEASFQNDLSQL--MLPDGLEKTESBAFTG--NPGDDHYNNRVVLWTGSKPKSPGLATE	385	
Db	1850	AEAKALQAMQALENSIOQOOTESGSKFINEKPKQDAQAAV	-----QNAKDLNQ	1902
Qy	386	NTYVNPKSLWQESPEIDYTKWLBEDTYQKNSVTGSPNKGLOKVKKNK	-----434	
Db	1903	TGNPTDKSQVEQ	-----LTQAVTTAKDNL-----HGDQKLARDQOQAVTTVNALP	1948
Qy	435	NLBHPKO	-----HNGVTITIGDNAPRVDFONKTLRKVLDIEVKLPSTIRKIGAFAP	487
Db	1949	NLMJAQOALTDAINAAPTREVAQHV	-----QTATELDHAMETLK--NKVDQVNTDKA	2000
Qy	488	QSNMLKSFEASDDLEEIKEGAF	-----MNNRI	514
Db	2001	QPNYT	---EASTDKKEAVDQALQAAESIITDPTNGSNANKDAVDQVLTKQEKENELNGE	2057
Qy	515	ETHELKDLVTIGDAAAPHINHIYAILVPESVQETGRSAFRONGANNLIFMGSKVKTIGEM	574	
Db	2058	RVAEAKTQAKQITDOLTHLN	-----ADQIATA--KQN-----IDQATKLQPIAEL	2100
Qy	575	AFLSNRL	-----EHLDDLSEQQLTEIPVQAFSD--NALKEVLLPASPALKTIREE	620
Db	2101	VDQATQLNQSDMLQQAQVNEHANVEQTVDT	-----QADSKQNAKQAIADA-----EN	2150
Qy	621	AFKQHLKQLEVASALSHIAPNALDDNGDBEQFNKVVVYKTH--H-----NSYALADGEH	673	
Db	2151	VLKQNAKQ-QVDQALQNI-LNNAQALNGDERV--ALAKTNGKHDIDQLNALNNAQDQ	2205	
Qy	674	F	-----IVDP-----DKLSSTIVDL	688
Db	2206	FKGRIDGSDNLNQIQIVDEKALNRAMDQLSQBITDNEGRTKSTNYVNADTQVKQYVD	2265	
Qy	689	EKILKIEGLDYSLRQTTOTQ	---FRDMTTAGKALLSKSNLRQGEKQKFIQ	737
Db	2266	ETVDKAKALDKSTGQNLTAQVILKNDADVTAAKKALNGEERLNNRKAELQRDLQTLH	2325	
Qy	738	---BAQFPLGRVLDKIAKAEKALVTTKATNGQLL	---ERSINKAVLA	781
Db	2326	NNAQRLAIQQIINNAETLNKASRAI--NRATKLDNMGVSQOQYIDEQHLGVISSTNYINA	2383	

Qy	782	-----YNN-----	SAIKKANVRKLEKELDLTLGLVEGKGPLAQATWVQG	820
Db	2384	DDNLKANYDNAIAAIAAHELDKVGQNAIAKAAEQIKQNIIDAQNALNGDQNLANAK	2439	
Qy	821	VYLLKTPPLPEYYVIGLVNVPFDKSGKLIYALDMSDTIGEGOKDAYGNPILNVDE	874	
Db	2440	-----DKNAFVNSL-----	2475	
Qy	875	-----DNEGYPHALAVATLADYEGLD-----	902	
Db	2476	IVNNQIDLDNDAETLKHVLVNEIPNA--BQTVNYQNADDNNAKTNFPDDAKRLANTLNSD	2532	
Qy	903	LSQLTSTR---QVPTAAYHRAGIFQAIONAAAEABQLLPKPGTHSEKSSSESANSKORG	959	
Db	2533	NTNVNDINGAIOAVNDIAHNLNGDQRLQDAKAIQSIQALANKLKEIEASNATDQKL	2592	
Qy	960	LQSN 963		
Db	2593	IAKN 2596		
RESULT 29				
ABM73008				
ID	ABM73008	standard; protein; 9535 AA.		
XX	AC	ABM73008;		
XX	AC			
DT	20-NOV-2003	(first entry)		
XX	DE	Staphylococcus aureus protein #2248.		
XX	KW	Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;		
KW	KW	enzymatic assay; antibiotic target.		
XX	OS	Staphylococcus aureus.		
XX	FN	W0200294868-A2.		
XX	PD	28-NOV-2002.		
XX	PF	27-MAR-2002; 2002WO-IB002637.		
XX	PR	27-MAR-2001; 2001GB-00007661.		
XX	PA	(CHIR-) CHIRON SPA.		
XX	FI	Masignani V, Mora M, Scarselli M;		
XX	DR	WPI; 2003-120786/11.		
XX	N-PSDB;	ACF74568.		
PT	PT	New Staphylococcus aureus protein, useful as a vaccine for treating or		
PT	PT	preventing Staphylococcal infection, specifically an infection caused by		
XX	XX	S. aureus, e.g. sepsis.		
PS	PS	Claim 1; SEQ ID NO 4496; 49pp; English.		
XX	CC	The invention relates to novel genes and encoded proteins from		
CC	CC	Staphylococcus aureus. A composition comprising the S. aureus protein, a		
CC	CC	nucleic acid encoding the protein, or an antibody to the protein, is		
CC	CC	useful as a pharmaceutical, particularly as a vaccine for treating or		
CC	CC	preventing infection due to Staphylococcus bacteria, specifically an		
CC	CC	infection caused by S. aureus. The composition is particularly useful for		
CC	CC	treating or preventing sepsis in a patient. The composition can also be		
CC	CC	used for diagnostics. The protein is also used in an assay for enzymatic		
CC	CC	studies and as a target for antibiotics. This sequence represents one of		
CC	CC	the novel S. aureus proteins of the invention		
SQ	SQ	Sequence 9535 AA;		

Query Match 4.1%; Score 207; DB 6; Length 9535;  
 Best Local Similarity 19.2%; Pred. No. 0.021;  
 Matches 231; Conservative 169; Mismatches 424; Indels 380; Gaps 52;





```
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2437 AA;

Query Match
Best Local Similarity 4.1%; Score 206.5; DB 4; Length 2437;
Matches 220; Conservative 157; Mismatches 420; Indels 361; Gaps 48;

QY 46 DVAESGSKLKINETSQVDTVDLFDSDKRTTPEKIKDNIAKAPROELKAVTENTES 105
DB 543 DHAKQTVSQAHLNNAQKHMETLI-----DSETRTAVKQDITAEQALDQALMDALQOQIA 598
QY 106 EKQITSGSQLEQSKESLSINKTPVSTSNWEICDFITKGNLTGLVSKSGVEK---LSQTDH 162
DB 599 DKDATRASAVYNAEPNKKQSVDEAVQNAE-----SIAGLNPTINKGNVSSATQA 650
QY 163 LVLPQAADGTOLI-----QVASFA-----FTPDKKTAAEYTSRAGENGESQLDVGK 212
DB 651 VISSKNALDGVVERLAQDKQTAGNSLNHLDTLTPAQQALENQINNATTRGEVAQKLTEAQ 710
QY 213 EIINEGEVENSILLKKVPIPTYKHIGQD-----AFVDNKN---IAEVNLP----- 255
DB 711 ALNQAMEALRNSIQDQOQTEAGSKFINKEDPKQDAYQAQVQNAKDLINQTNPTLDKAQV 770
QY 256 BSETISDYAPAHALAKOIDLDPNLKAIGELAFFDNOITGKLSLPQLMRLAERAFKSNH 315
DB 771 EQLTAQVNAQKLNHGDKLADDKQHAVTDL-----NQLNGLNPNQOAL-----ESQ 818
QY 316 IKTIIFRGNLSKVIGEASFQDNDLSQL--MLPDGLEKIESEAFPG--NPGDDHYNNRVYL 371
DB 819 INNAATRGVEVAKLAERAKALDQAMQALRNSIQDQOQTESGSKFINKEDPKQDAYQAQV-- 876
QY 372 WTKSGKNPSGLATENTYVNPDKSLMQESPEIDYTKWLEDFTYQKNQSVTGSNKGLOKVK 431
DB 877 ----QNAKDLINQTNPTLDKQSVQEQ-----LTQAVTTAKDNL-----HGDKLA 917
QY 432 RNK-----NLEIPKQ-----HNGVTITEIGDNAFRNVDQFNKTLRKYLDEEV 473
DB 918 RQQAQAVTVVNALPNLHNAQQAALTDINAATRTVEVAQH-----QTATELDHAMEYL 971
QY 474 KLPSTIRKIGAPAFOSNNLKSPEASDDEEIEKEGAF-----EHLDLSEKQLTEIPQAFSD--NALK 606
DB 972 K--NKVDQVNTDKAPNYT--EASTDKKEAVDQALQAAESITDPTNGSNANKDAVDQVL 1026
QY 510 -----MNRRIETLEKOKLVIGDMAFHINHIYAILVPESVQEBIGRSAPFRONGANN 560
DB 1027 TKLQEBENELNGERNVAEAKTQAKOTIDQTLN-----ADQIATA--KQN----- 1070
QY 561 LIPMGSKVKTGEMAFSLNRL-----EHLDLSEKQLTEIPQAFSD--NALK 606
DB 1071 -IDQATKQPIAELVDQATQLNQSDQDQLOQAVNEHANVEQTVDYT-----QADSKQNAVK 1125
QY 607 EYLLPASLKTIRREAFKQKHLKQLEVASALSHIAFNALDNDGDQFNKVVVVKTH--H- 663
DB 1126 QAIADA-----ENVLKQNAKQ-QVDQALQNI-LNAKQALNGDSRV--ALAKTNGKHD 1174
QY 664 ----NSYALADGEHP-----IVDP-----DKLSSSTIVL----- 688
DB 1175 IDQLNALNNAQDQFGKGRIDQSDNLDNLQIQIIVDEAKALNRAVDQISQETIDNEGRTKGT 1234
QY 689 -----EKILKIEGLDYSLRQTTQ-----FRDMTATGKALLSKSNLRQGE 731
DB 1235 NYVNADTVQKVYDETVDKAKQALQKSTQGNITAKQVINKLNDAVTAAKALNGEBRLNR 1294
QY 732 KQKFLQ-----EAQFFLGRVLDLKAIAKAEKALVTTKATNGQGLL----BRSKN 777
DB 1295 KAEALQRLDQLTHLNAQQLAIQIINNAETLNKASRAI--NRATKLDNAMGAVQYIDE 1352
QY 778 AVLA-----YNN-----SAIKKANVKLEKELDLTLGLV 806
DB 1353 OHLGVISSTNYINADDNLKANTDNIANAHAELDKVQGNIAKAEABQLKQNIIDAQNAL 1412
QY 807 ECKGPLAQATMVGQVYLLKTPPLPEYIIGLVNVPDKSGKLIYALDMSDTIGEGQKDAYG 866
```

---

```
DB 1413 NGDQNLAK-----DKANAFVNSL---NGLNQOQDLAH 1444
QY 867 NPILNVDE-----DNEGTHALAVATLADYEGLD----- 894
DB 1445 KAINNADTVSDVTDIVNNQIDLNDAMETLKLHVDNEIPNA---EQTVNYQVADNNAKTNP 1501
QY 895 -----IKTILNSKLSQTSIR---QVPTAAYHRAGIFQALQNAAAAEQLLPKPGTHSE 945
DB 1502 DDAKRLANTLNSONTNVNDINGAIQAVNDAIHNLNGDRLQDAKDAKAIOSINQALANKL 1561
QY 946 KSSSESANSKDRGLQSN 963
DB 1562 KEIEASNATDQDKLIAXN 1579

RESULT 32
AAU37403
ID AAU37403 standard; protein; 6281 AA.
XX AC AAU37403;
XX DT 14-FEB-2002 (first entry)
XX Staphylococcus aureus cellular proliferation protein #1573.
DE Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX Staphylococcus aureus.
FN WO200170955-A2.
PD 27-SEP-2001.
XX 21-MAR-2001; 2001WO-US009180.
XX 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207272P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX (ELIT-) ELITRA PHARM INC.
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
DR N-PSDB; AAS55262.
XX New polynucleotides for the identification and development of
antibiotics, comprise sequences of antisense nucleic acids.
XX Example 3; SEQ ID NO 12996; 511pp; English.
CC The invention relates to antisense inhibitors of genes essential to
prokaryotic cellular proliferation, their use in identifying the genes,
their use in the discovery of novel antibiotics, the essential genes,
themselves and the encoded proteins. The prokaryotes used are Escherichia
coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
useful for the identification of potential new targets for antibiotic
development. The antisense nucleic acids can also be used to identify
proteins used in proliferation, to express these proteins, and to obtain
antibodies capable of binding to the expressed proteins. The proteins can
be used to screen compounds in rational drug discovery programmes. The
antisense nucleic acid sequence is also useful to screen for homologous
nucleic acids which are required for cell proliferation in a wide variety
of organisms. The present sequence represents an essential prokaryotic
cellular proliferation protein. Note: The sequence data for this patent
```



CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 6281 AA;

```
Query Match          4.1%; Score 206.5; DB 4; Length 6281;
Best Local Similarity 19.0%; Pred. No. 0.013;
Matches 220; Conservative 157; Mismatches 420; Indels 361; Gaps 48;

QY 46 DVAESSGSKLKNETSGPVDVDTLFDKRTTPEKIKDNLAKGPREBELKAVTENTES 105
DB 4385 DHAKQTVSOLAHNNAQKMHEDTLI-----DSETRTAVKQDLTEAQAQLDQALDQOOSTA 4440
QY 106 EKQITSGSQLESKESLSLNKTPVSTSNWEICDPITKGNLTVLGSKSGVEK---LSOTDH 162
DB 4441 DKDATRASSAYVNAEPNKKQSYDEAVQNAE-----SIIAGLNPTINKGNVSSATO 4492
QY 163 LVLPSQAADGTQLI-----QVASPA-----FTPDKKTAAJAYTSRAGENSEISQLDVGDK 212
DB 4493 VISSKNALDGVRLAQDKQTAGNSLNHLDLQTLTPAQOQALENQINNATTRGEVAQKLTEAQ 4552
QY 213 EIINEGEVFNYSLLKKVYTPYTKHIGOD-----AFVDNKN--IAEVNLP-----255
DB 4553 ALNQAMEALRNSIQDQOQTEAGSKFINEDKPKDAYQAAVQNAKOLINQTNPTLTDKAQV 4612
QY 256 ESLETTISDVAFALHALKQIDLPNKAIGELAFDQITGKLSLPRQLMLRAERAPKSNH 315
DB 4613 EQLTQAVNQAKNHLGDKLADKQHAVTDL-----NQLNGLNPPQOAL-----ESQ 4660
QY 316 IKTIIEFRGNSLVIGESFQDNDLSQL--MLPDGLEKIESEAPTG--NFGDDHNNRVVL 371
DB 4661 INNAATRGEVAQKLAERAKALDQAMQALRNSIQDQOQTESSGFNEDKPKDAYQAAV-- 4718
QY 372 WTKSGKNPSGLATENTYVNPDKLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVK 431
DB 4719 -----QNAKDLINQTNPTLDRKSQVEQ-----LTOAVTTAKDNL-----HGDQKLA 4759
QY 432 RNK-----NLEIPKO-----HNGVTITEIGNAFNPNVDNFQNKLRKYDLEEV 473
DB 4760 RQOQAVTTVNALPNLHNAHQOQALTDAINAAPTRTEVAQHV-----QTATELDHAMETL 4813
QY 474 KLPSTIRKIGAPAFQSNLKSFEASDDLEEIEGAF-----DPTNGSNANKDAVDQVL 4868
DB 4814 K--NKVDQVNTDKAQNVT--EASTDKKEAVDQALQAAESIPTDPTNGSNANKDAVDQVL 4868
QY 510 -----MNNRIETLEKDKLVITGDAAFHINHIYALVPESVQEIQRSPFRONGANN 560
DB 4869 TKLQEKENELNGNERNVAEAKTQAKOTIDQTLHLN-----ADQIATA--KQN-----4912
QY 561 LIFMGSKVKTGEMAFLSNRL-----EHLDSLSEQKOLTEIPVQAFSD--NALK 606
DB 4913 -IDQATKQPIAEVLVDQATQLNQSMDQOLQAVNEHANVEQTVDYT-----QADSQKQNAVY 4967
QY 607 EVLLPASLTIRIEAPFKQKHLKQEVASALSHIAFNALDNDQDEQFNDKVVVVKTH--H- 663
DB 4968 QAIADA-----ENVLKQNAKQ-QVDQALQNI-LNAKQALNGDERV--ALAKTNGKHD 5016
QY 664 ----NRYALADGHP-----IVDP-----DKJSSSTIVDL-----688
DB 5017 IDQLNALNNAQDGFGRIDQSDNLDNQIQIIVDEAKALNRAMDQLSQEITDNEGRKTGST 5076
QY 689 -----EKILKILEGLDYSLRQTQTQ-----FRDWTAGKALLSKSNLRQGE 731
DB 5077 NYNADTQVKQVYDETVDKAKQALDKSTGQNLTAQKVIKLNDAVTAAKKALNGEERLNR 5136
QY 732 KQKFLQ-----EAGQFLGRVLDLKAIAEKALKVTKATKNGQLL---ERSINK 777
DB 5137 KAEALQRLDQLTHLNAQRLQALQOINNAETLNKASRAI--NRATKLDNAMGAVQOYIDE 5194
QY 778 AVLA-----YNN-----SAIKKANYKRLKLELDLITGLV 806
DB 5195 QHLGVISTNYINADDNLKANYDNIANAHAHELDKVQGNALIAKAEQKQNIIDAQNAL 5254
```

```
QY 807 EKGGLPLAQAATMVQGVLLKTLPLPEYVYIGLVNVPYFDKSGKLIYALDMSDTIGEGOKDAYG 866
DB 5255 NGDQNLANAK-----DKANAFVNSL---NGLNQOQODLAH 5286
QY 867 NPLNVDE-----DNEGYPHALAVATLADYEGLD-----894
DB 5287 KAINNADTVSDVTDIVNNQIDLNDAMETLKLHVDNEIPNA---EQVTNYQNADNNAKTNF 5343
QY 895 -----IKTILNSKLSQTSIR---QVPTAAVHRAGIFQAIQNAIAAEAEQLLPKPTHSE 945
DB 5344 DDAKLANLTLLSNDNTNVNDINGAIOAVNDAIHLNGLQRLQDAKDKATQISINOALANKL 5403
QY 946 KSSSESANSKDRGLQSN 963
DB 5404 KXIEASNATDQDKLIAKN 5421

RESULT 33
ABJ19119
ID ABJ19119 standard; protein; 10498 AA.
XX AC ABJ19119;
XX DT 06-MAR-2003 (first entry)
XX DE Pathogen specific antigen related staphylococcal protein SEQ ID No 440.
XX KW Antibacterial; virucide; fungicide; protozoicide; cytostatic; anti-HIV;
hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
XX OS auto-immune disease; HIV; hepatitis.
XX OS Staphylococcus sp.
XX PN WO200259148-A2.
XX PD 01-AUG-2002.
XX PF 21-JAN-2002; 2002WO-EP000546.
XX PR 26-JAN-2001; 2001AT-00000130.
XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX PI Meinke A, Nagy E, Von Ahsen U, Klade C, Henics T, Zauner W;
PI Minh DB, Vycvytska O, Etz H, Dryla A, Weichhart T, Hafner M;
PI Tempelmaier B;
XX DR WPI; 2003-075410/07.
XX PT Identifying, isolating and producing hyperimmune serum-reactive antigens
from a pathogen, for preparing vaccine or medicament for treating or
preventing e.g. staphylococcal infections, comprises providing antibody
preparation.
XX PS Example 7; Page 234-235; 252pp; English.
XX CC The invention relates to a novel method for identifying, isolating and
producing hyperimmune serum-reactive antigens from a pathogen, tumour,
allergen, a tissue or host prone to auto-immunity, where the antigens are
used in a vaccine, comprises providing antibody preparation from a plasma
pool of a type of animal, or individual sera with antibodies against the
specific pathogen, tumour, allergen, tissue or host prone to auto-
immunity. The hyperimmune serum-reactive antigens comprising any of the
62 sequences of 53-2261 amino acids fully defined in the specification,
or their hyperimmune fragments are useful for the manufacture of a
pharmaceutical preparation, particularly a vaccine against staphylococcal
infections or colonisation against S. aureus or S. epidermidis. The
preparation of antibodies is useful for the manufacture of a medicament
for treating or preventing staphylococcal infections or colonisation
against S. aureus or S. epidermidis. The antibody preparations may also
be used for diagnostic and imaging purposes. Other conditions that can be
```





[illegible]

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC [fip.wipo.int/pub/published/pct/sequences](http://fip.wipo.int/pub/published/pct/sequences)

Sequence 6641 AA;

Query Match 4.0%; Score 205; DB 6; Length 6641;  
Best Local Similarity 19.1%; Pred. No. 0.017;  
Matches 201; Conservative 178; Mismatches 447; Indels 224; Gaps 45;

Qy	47	YA8SSGSKUKINETSPV---DDTVTDLFSKRTTPPKIKONLAKGPRBOELKAVTENTE	104
Db	4050	YNQAVKAKNIINDQPTVPVNADEIQVLVEKQT-----KDNL-----HGDQKLANDKTD	4100
Qy	105	SEKQITSGSOLSESKESLSLNKTVPSNWEICDFTKGNLTVLGLSKSGVEKLSQTDHLV	164
Db	4101	AQATLNALNYLNOAQRNLKTKVQNSRSPREVQKVQLANQNLNDAMKCLDALTGDAIK	4160
Qy	165	LPQAAOGTOLIQVAFPTDKKTAIAEYTSRAGENGESQL-----DVGKKEII	215
Db	4161	QTSNYINEDTSQQVDFEYDRGKNIIVAEOTNPNMSPNTINTADKITEAKNDLHGVOKL	4220
Qy	216	NEGVEFNSYLLKKVT-IPTGVGHIGQDAPVDNKNIAEVN-----LPESLETISDYAF	266
Db	4221	KQAQQQINTINGLNQAQKEQLNOEIQOTTRSEVHVINKAQLNDMNTLQKSIT	4280
Qy	267	AHLAKQIDLPNLKAIGELAFPDN-----QITGKLSLPRQLMLRAERAFKSNHKTIE	320
Db	4281	DEHEVKQTSYIN-ETVGNQTAYNADVQVKQIINQTSNPTMNPLEVERA-----TSNVK	4334
Qy	321	FRGNSLKVIGEASQPDNDLSQLMLP---DGLKXIESEAPTG-----NPGDDHYNRV	369
Db	4335	ISKDALH--GERELNDKNKSKTFAVNLNDLNOAQKEALTHEIEQATIVSQVNNIYNKAK	4392
Qy	370	VLMTKSGK-----NPSGLATENTYVNPD---KSLMOE-----SPEIDYTKW	407
Db	4393	ALANDMKKLDIVAQQDNVRQSNYINEDSTPQNMVYNDTINHAQSIIDQVANPTMSHDE-	4451
Qy	408	LEEDFTQKNSVTGFSNK-GLQKVRKNKLEIPKHQNGVITTEIGDNAPRVDPONKTLR	466
Db	4452	IENAINNIKHAINALDGEHLQQAKEANVL-----LINSLNDLNAQPDARINLVN	4502
Qy	467	KYDLEEVKLPTIRKIGAPAFQSNLKSFEASDDLEBIKEGAFMNNRIETLELKDCLVTI	536
Db	4503	EAQTR-----KV-----AEQLQSAQALNDAMKHLRNSIQN---SSVRQESKYINA	4546
Qy	527	GDA-AFHINHIYAI-----VLPESVQBIGRSAPRQ-----NGANNLFMGSKVKTIGEMA	575
Db	4547	SDAKKEQYNH-AVREVENIINEHQPTLKDKEIIKQLTDGVNQANNDL-----	4591
Qy	576	FLSNRLEHLDLSEQKQLTEIP---VQAFSDNALKE-----VLLPAS	613
Db	4592	---NGVELLDADKQNAHQSIPTLMHLNQAQNALNEKINNNAVTRTEVAALIGQAKLLDHA	4648
Qy	614	LKTIREEAFKKNHLKQLEVASALSIAFNAL-DDNDGDEQFDN-----KVWVKTHNSY	666
Db	4649	MENLEESI KDKBQVKQ-----SSNYINEDSDVQETYDNAVHDVHTILNQVNP TL	4698
Qy	667	ALADGEHFIYDVKLSSTIVDLEKILKLEGLD--YSTLRQTTQTQTFRDMTTAGKALLSK	724
Db	4699	SIEDIEHAINEVNAKKQLRGKQKYQTIIDLAKBELSKLDLTSQQSSSIISNOIYTAKTR	4758
Qy	725	SNLRQG-EKQFLQEAQFFLGRVDLDKAIKAERKALVTKKATKNQGLLERSINKAVLAYN	783
Db	4759	TEVAQALEKAKSLNHAM-----KALNKVYKNADKVLDSRFINEDQPEKKAYQQAINHVD	4813
Qy	784	NSAIKKAN-----VKRLEKELDLLTGLVEGKGPLAQ-----ATMVQGVYLLKTPPLP	831
Db	4814	SIIHROTNPENDPTVINSITHELETAQNKLHGDKQLAHQAQDARNVING-LHLHNAVQR	4871



```
Db 7729 QTSNVINEDTSQVNFDEYTBGKNVAVQINPNPNTNINTIADKI TEAKNDLHGQKL 7788
Qy 216 NEGEVFNLYLKVVT-IPYGYKHIGQDAFVNKNIAFVN-----LPSELETISDYAF 266
Db 7789 KQAQQOSINTINQMTGLNQAQKEQLNQBIEIQTOTRSEVHQVINKAQAALNDSNVTLRQSIT 7848
Qy 267 AHLALKQDLDLNDLKAIGELAPFDN-----QITGKLSLPQLMLAERAFKSHIKTIE 320
Db 7849 DEHEVYQTSNYIN-ETVGNQTAYNNAVDVRVQIINQTSNPTMNPLEVERA-----TSNVK 7902
Qy 321 PRGNSLKVIGEASFODNDLSQMLP---DGLEKIESEAPTG-----NPGDDHYNRV 369
Db 7903 ISKDALH--GERELNDNKNSTFPAVNHLDNLNQAQKEALTHEIQATVTSQVNNIYNKAK 7960
Qy 370 VLMTKSGK-----NPSGLATENTYVNPD---KSLMQE-----SPEIDYTKW 407
Db 7961 ALNNDMKKLKDIVAQODNVRSNNYINEDSTPQNNYNDTINHAQSIIDQVANPTMSHDE- 8019
Qy 408 LEEDFTYQKNSVTGFSNK-GLQKVRGNKLEIPKOHNGVTTIEIGDNAFRNVDFQNKTLR 466
Db 8020 TENAINNIKHAINALDGEHLQQAQENANL-----LINSLDNLNAPQORDAINRLVN 8070
Qy 467 KYDLEEVKLPSTIRKIGAFAPFOSNNLKSFEASDLEETKEGAFMNNRIETLELKDCLVTI 526
Db 8071 EAQTRÉ-----KV-----AEQLQAQALNDAMKHLRNSIQNQ--SSVRQESKYINA 8114
Qy 527 GDA-AFHINHIYAI-----VLPEVQEIGRSAFRQ-----NGANNLIIFMGSKVKTJGEMA 575
Db 8115 SDAKKEQYNH--AVREVENIINEQHPITLDEIKITQITDGVNQANNDL-----8159
Qy 576 FLUNLEHLDESEQOLTEIP-----VOAFSDNALKE-----VLLPAS 613
Db 8160 ---NGVELLDADKQNAHOSIPTLHNLQAQNALNEKINNNAVTRTEVAAIIGQAQLDHA 8216
Qy 614 LNTIREAFKKNHLKQLEVASALSIAFNAL--DDNDGDBQFDPN-----KVVKTHNSY 666
Db 8217 MENLESIKDKEQVQK-----SSNYINEDSDVQETVDYNDVHTBIUNQVNTPL 8266
Qy 667 ALADGEHTVDPKLSSTIVDLEKILKLEGLD--YSTLRQTTQTQFRDMTTAGKALLSK 724
Db 8267 STIEDIEHAINEVNAQKQLRGQKLYQITDLADKELSKLDLTSQSSSISNQIYTAKT 8326
Qy 725 SNLRGQ-EKQKPLQBAQFPLGRVDLDKAIAKAEKALVTKKATYNGQLLERSINKAVLAYN 783
Db 8327 TEVAQAIRKAKSLNHAM-----KALNKVYKNADKVLDSRRFINEDQPEKAYQQAINHVD 8381
Qy 784 NSAIKKAN-----VKLEKELDLTLGLVEKGKPLAQ-----ATMVGVVLLKTPPLP 831
Db 8382 SIIRHQTNPEDMPTVINSTHELETAQNHLHGDQQLAHAAQQAANVING--LIHLNVAQR 8439
Qy 832 EYTGGLNVYFDKSGKLIYALDMSDTIGEC-----QKDAYGNPILN-----VQEDNEGYHAL 882
Db 8440 EVMINTNTNATREKAVKNDNAQALDKAMETLQQVVAHKNNILNDSKYLNEDSK-YQQQ 8498
Qy 883 AVATADVEGLDKITILNSKLSQLSIRQVPTAAVHRAGIFQAIQNAABAEQLLPKPGT 942
Db 8499 YDRIADAEQL-----LNQTTN-----PTLEPYKVDIVK--DNVLANEKILF-----8538
Qy 943 HSEKSSSESANSDRGLOSNPKTRGRHS 972
Db 8539 GAELKYDYSK-NANDEIKHMYLNNAQKQS 8567
RESULT 38
ID ADS04803
AC ADS04803
XX ADS04803;
DT 04-NOV-2004 (first entry)
DE Staphylococcus epidermis polypeptide seqid 4098.
```

```
XX
KW antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
KW recombinant expression vector; infection; computer readable medium;
XX computer based system.
OS Staphylococcus epidermidis.
XX US2004147734-A1.
XX 29-JUL-2004.
XX 01-DEC-2003; 2003US-00724972.
XX 08-NOV-1997; 97US-0064964P.
XX 13-AUG-1998; 98US-00134001.
XX 29-NOV-1999; 99US-00450969.
XX (DOUC/) DOUCETTE-STAMM L.
XX (BUSH/) BUSH D.
XX Doucette-Stamm L, Bush D;
XX WPI; 2004-580138/56.
XX N-PSDB; ADS01031.
XX New isolated polypeptide and encoding nucleic acid derived from
XX Staphylococcus epidermidis, useful for diagnosing, preventing and/or
XX treating an S. epidermidis bacterial infection.
XX Claim 17; SEQ ID NO 4098; 741pp; English.
XX The invention describes an isolated nucleic acid comprising a nucleotide
XX sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
XX 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
XX of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
XX given in the specification. Also described are: a recombinant expression
XX vector; a cell comprising a recombinant expression vector of (1);
XX producing an S. epidermidis polypeptide; an isolated nucleic acid
XX comprising a nucleotide sequence of at least 8 nucleotides in length; a
XX vaccine composition for prevention or treatment of an S. epidermidis
XX infection, comprising a nucleic acid cited above and a carrier; treating
XX a subject for S. epidermidis infection; a recombinant or substantially
XX pure preparation of an S. epidermidis polypeptide or its fragment; a
XX vaccine composition for prevention or treatment of an S. epidermidis
XX infection; detecting the presence of a Staphylococcus nucleic acid in a
XX sample; a computer readable medium having recorded in it the nucleotide
XX sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
XX system for identifying fragments of the Staphylococcus genome of
XX commercial importance; a computer based system for identifying fragments
XX of the Staphylococcus plasmids of commercial importance; identifying
XX commercially important nucleic acid fragments of the Staphylococcus
XX genome and/or plasmids; and identifying an expression modulating fragment
XX of the Staphylococcus genome and/or plasmids. The methods and
XX compositions of the present invention are useful for the diagnosis,
XX prevention and/or treatment of an Staphylococcus epidermidis bacterial
XX infection. This is the amino acid sequence of a S. epidermis protein of
XX the invention.
XX
SQ Sequence 10203 AA;
Query Match 4.0%; Score 205; DB 8; Length 10203;
Best Local Similarity 19.1%; Pred. No. 0.031;
Matches 201; Conservative 178; Mismatches 447; Indels 224; Gaps 45;
Qy 47 YAESGSKSKLINETSQGV--DDTVTLDFSDKRTTPEKIKDNLAKGPREQELKAVTENTE 104
Db 7618 YNQAVNKAKNIINDQPTFVMADEIQSVLNEVKGT----KQNL-----HGQKLANDKTD 7668
Qy 105 SEKQITSGSQLEQSKESLSLNKTVPTSTNWEICDPTKGNITLVGLSKSGVEKLSQTHLV 164
Db 7669 AQATLNALNYLNQAQGNLETKVQNSNSRPEVQVQLANQLNDAMKLDLALTGNDAIK 7728
Qy 165 LPSQAADGTQLIQVASFAPFTDKKTAETYSRAGEISQL-----DVGDKELI 215
```

```
Db 7729 QTSNYINEDTSQVNFDEYTRDGRKNIQAETQPNMSPNTINTIADKITEAKNDLHGQKL 7788
Qy 216 NEGEVFNSTLLKKVFT-IPGYKHIGQDAFVDNKNIAEVN-----LPESLETISDYAP 266
Db 7789 KQAQOQSINTINQMTGLNQAQKEQLNOEIQOTTRSEVHOVINKAQAALNDSNLRQSIT 7848
Qy 267 AHLALKQIDLPNLKAIGELAFDNL-----QITGKLSLPRLMLRAERAFKSNHIKTIE 320
Db 7849 DEBEVKQTSNYIN-ETVGNQATYNNADVVRKQIINQTSNPTWNPLEVERA-----TSNVK 7902
Qy 321 PRGSLKVTIGEASFQNDLSQLMLP-----DGLEKIESEAFG-----NPGDDHYNRV 369
Db 7903 ISKDALH--GERELNDKNKSTFAVNHLDNLNQAQKEALTHEIQATIVSQVNNIYNKAK 7960
Qy 370 VLWTKSGK-----NPSGLATENTYVNP-----KSLWQE-----SPEIDYTKW 407
Db 7961 ALNNDMKKLKDIVAQODVVRQSNVYNEDSTQNNINDTINIAHQIIIDQVAMPTWSDHE- 8019
Qy 408 LEEDFTYQKNSVTFNSK-GLQVKRNKNLEIPKQHGVTITEIGDNAPRNVDFOKTLR 466
Db 8020 IENAINNIKHAINDLGEHKLQAQKENANL-----LINSLDLNAQORDAINRLVN 8070
Qy 467 KYDLEVKLPSTIRKIGAPAPOSNNLKSPESADDEEIKEGAFMNRRIETLEKDKLVTI 526
Db 8071 EAQTRE-----KV-----AEQLQSAQALNDAMKHLRNSIQNQ--SSVRQESKYINA 8114
Qy 527 GDA-AFHINHIYAI-----VLPESVQETGRSAFRQ-----NGANNLIPMGSKVKTLGEMA 575
Db 8115 SDAKKEQVNH--AVREVENIINEQHTLDEKIIKQITDGVNQANNDL-----8159
Qy 576 FLNRLHLDLSEQQLTEIP-----VQAFSDNALKE-----VLLPAS 613
Db 8160 ---NGVELLDADQNAHQHSIPTLMLHNOAQNALNEKINNNAVTRVEAAIIGQAKLLDHA 8216
Qy 614 LKTIREEAFKQHLKQLEVASALSHIAFNAL--DDNDGDEQFDN-----KVVVTHHNSY 666
Db 8217 MENLEESIKDKQVQKQ-----SSNYINEDSDVOETDYNADVHVTEILNQTVNPTL 8266
Qy 667 ALADGEHFIVDPDKLSSITVDLEKILKLEGLD--YSTLRQTOTQTFQFRDMMTAGKALLSK 724
Db 8267 SIEDIEHAINEVNQAKQLRGKQKQYQITDIDLADKELSKLDDLTSSQSSISNQIYAKTR 8326
Qy 725 SNLRQG-EKQKPLQBAQFPLGRVLDKATAKAELVTKKATKNGQLLERSINKAVLAYN 783
Db 8327 TEVAQAEKAKSLNHAM-----KALNKVYKNADKVLDSRFNEDQPEKKAQQAINHVD 8381
Qy 784 NSAIKAN-----VKLEKELDLLTGLVEGKGLAQ-----ATWQGVYLLKTLPLP 831
Db 8382 SIIHRQTNPEMDPTVINSTHETEAQNNLHGDQKLAHAQQAANVING--LIHLNVAQR 8439
Qy 832 EYIIGLVYFDKSGKLIYALDMSDTIGEG-----QKDAGNFIIN-----VDEDNRYHAL 882
Db 8440 EYVINTNTNATREKAVKNDNAQALDKAMETLQOVAAKHNNILNDSKYLNEDSK-YQQQ 8498
Qy 883 AVATLADVEGLDKILNLSKLSQTSIRQVPTAAVHRAGIFQAIQNAAAEAEOPLLKPGT 942
Db 8499 YDRVIADAEQL-----LNQTN-----PTLEPKYDIVK--DNVLANEKILP-----8538
Qy 943 HSEKSSSESANSKDRGLQSNPKTNRRGHS 972
Db 8539 GAELSYDKS-NANDEIKHMVNLNNAQKQS 8567
```

## RESULT 39

ABU15838

ID ABU15838 standard; protein; 2481 AA.

XX

AC ABU15838;

XX

DT 19-JUN-2003 (first entry)

XX

DE Protein encoded by Prokaryotic essential gene #1365.

```
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Staphylococcus aureus.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA19708.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 43762; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 2481 AA;
```

Query Match 4.0%; Score 204.5; DB 6; Length 2481;

Best Local Similarity 18.5%; Pred. No. 0.0046;

Matches 206; Conservative 159; Mismatches 403; Indels 343; Gaps 45;

Qy 10 LTLTTSVVTTHNQEVFSLVKEPILKQ-----TQASSSI-----SGADYASSGSKSKKI 58

Db 1293 LQATQTNVINNDQNAETNEEKAIIQQLATAVTADKNNITAAATDNDGVDGTAKDAGNSIQS 1352





[illegible]

RESULT 41

ABU15887  
ID ABU15887 standard; protein; 6713 AA.

XX

AC ABU15887;

XX

DT 19-JUN-2003 (first entry)

XX

DE Protein encoded by Pro

[illegible]

```

Query Match      4.0%; Score 203; DB 6; Length 6713;
Best Local Similarity 19.5%; Pred. No. 0.023;
Matches 232; Conservative 183; Mismatches 396; Indels 376; Gaps 64;

          y       b
34 KQTQASSISGAD-----YAESSGSKLKLNETSGP-VDDTVTLDFSDKRTTPEIKDN 86
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
5050 QQTQAGSFXINEDKPQDAYQAQVNAKDLLNQTCNPTLDKSQEQLTQAVTT---AKDN 5106

          y       b
87 L---AKGPREQELKAVTN-----TSSEKOITSGSOLEQSKE 120
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
5107 LHGDQKLARDQQAAVTTVNALPNLNHAQQOQLTDAINAAPRTEVAHQVQTATELDAME 5166

          y       b
121 SLSLNKTVPSNSWEICDFTFKGNLTVLGSXSGVEKLSQTDHLVLPSQAADGTGLIQVAS 180
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

```

Db 5167 TLK-NKV-----DQVNTDKAQPNTYEASTDKKAVD-----QALQRAQ 5203  
QY 181 FAFTPD-----KKTAAEYTSRAGNEISQDQVDGKEIINEGEVFNYSLLKVVPTPGY 235  
Db 5204 SITDPTNGSNANKDAVEQAALTKLB--KVNEL--NGNERVAEAKTQAKQTDQLT----- 5254  
QY 236 KHIGQDAFVDKN-----IAEV-----NLPESLETISDYAFALHALKQ-IDL----- 276  
Db 5255 -HLNADQIATAKONIDQATKLPQIAELVDQATQNLQMSDQQLQAAVNEHANVEQTIDYTOA 5313  
QY 277 -PDNLKAIGE-LAFPDN--QITGKLSLPLROLML--AERAP-----KSNHIKIE 320  
Db 5314 DSDQKQAYQAADAEVNLKQWANKQVQDQALONTLNKAKQALNGDERVALAKTNGKHID 5373  
QY 321 FRGSLKVIGEASP-----QNDLSQLM-----LPDGLKIESEAFPTGNPGDDHYN 367  
Db 5374 -QLNALNNAQQDGFGRIGDQSNLQIQIIVDEAKALNRAMDQL--SQEITGNEGR----- 5426  
QY 368 RVVLWTKSGKNPSGLATENTYVNPDKSLWQESPE-IDYTKW-----LE 409  
Db 5427 -----TKGSTN-----YVNADTVQKVYDEAVDKAKQALDKSSGQNLTAEQVILN 5472  
QY 410 EBFYTKNSVTGFS-----NKGLOKVR 432  
Db 5473 DAVTAAKKALNGEERLNKKBALQRLDQLTHLNAQRLAQIQINNAETLWASRAINR 5532  
QY 433 NKMLE-----IPKQHGNY-----TITEIGDNAPRV-----DFQNKTLRK 467  
Db 5533 ATKLNAMGAVQYIDEGHLGVISSTNYINADDNLKANYDNAIAAHELDKVCQNAIAK 5592  
QY 468 YLEEVK-----LPSTIRKIGAFQSNL-----KSFASDDLEB 503  
Db 5593 ABAEQKLQNIIDAQNALNGDQNLAKAKANAFVNSLNGLNQOQDQDLAKHAINNADTVSD 5652  
QY 504 IKEGAFMNRRIETLEKDKLVIGDAAFHINHYAIVLPESVQIEGRSAFRONGANNLIF 563  
Db 5653 VTD--IVNNQID-----LNDAMETLKHLDVNEIPNABQTVNYQVADNNAKTINFD 5700  
QY 564 MGSVKYTL-----GBMAFLSNRLEHLDLSEKQ-LTEIPVQAFSD---NALKE 607  
Db 5701 AKRLANTLNSDNTVNDINGAIAQVNDAIHNLGDQRLQDAKDAKAIQSIQNALANKLE 5760  
QY 608 VLLPASLTIREAPKQHLKQLEVASALSHTAFNALDNDGDEPD---NKVVVKTTHN 664  
Db 5761 I--EASNATDQDKLAKKAEL--ANSINNIT--NKATSNQAVSQVQTAGNHAIBQVHAN 5815  
QY 665 SYVALDGEHFIVDPDKLSSTIVDLK-ILKLTIEGLDYS-----TLRQTTQ 710  
Db 5816 EIPKAK-----IDANK-----DVKQVQALLIDEIDRPNLTDKQKALKRINQILOQ 5864  
QY 711 PRDMTTAGKALSKSNLRGGEKQ--KFLQEAQFFL-GRVDLDKAIKAEKALVTK----- 762  
Db 5865 HNDINNA---LTKEEIEQAQAQALQDIKLVKAKEDAKQDVQKQVQALIDEIDQNP 5920  
QY 763 -KATNGQLLERSINKAVLAYNNSAIKANVRLSKELDLLTGLVEGKGPLAQATWQGV 821  
Db 5921 NLTDREKQALKDRINQ--ILQOHHNGINNMTK---EIEE-----QAKAQAQA----- 5964  
QY 822 YLLKTPPLPEYVIGLVYFDKSGK-LIIVALDMSDTIGEGOKDAYGNPLINVDSD--NEG 878  
Db 5965 --LKEIKLVKAKENAKQDVQKQVQALIDEIDQNPNTDKEKQALKDRINQILOQCHNDI 6022  
QY 879 YHAL-----AVATLADYEGH-DIKTLNSKLSQTSIROQVPTAAYHRAGIFOIONAA 930  
Db 6023 NNAMTKEEIEQAQAQA--QALQDIKLVKAKEDAKNAIKAL-----AN 6064  
QY 931 ABAEQLLPKPGTHSEKSSSS--ESANSKDRGLQS--NPKT-----NRG 969  
Db 6065 AKRQDQINSNPDLTPQKAKALKEIDEAEKRALQNVENAQTTDQLNRG 6111

ID AU34207 standard; protein; 2025 AA.  
XX AC AAU34207;  
XX DT 14-FEB-2002 (first entry)  
XX DE Staphylococcus aureus cellular proliferation protein #483.  
XX KW Antisense; prokaryotic cellular proliferation protein; antibiotic;  
XX KW antibacterial; drug design.  
XX OS Staphylococcus aureus.  
XX PN WO200170955-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US009180.  
XX PR 21-MAR-2000; 2000US-0191078P.  
XX PR 23-MAY-2000; 2000US-0206848P.  
XX PR 26-MAY-2000; 2000US-0207727P.  
XX PR 23-OCT-2000; 2000US-0242578P.  
XX PR 27-NOV-2000; 2000US-0253625P.  
XX PR 22-DEC-2000; 2000US-0257931P.  
XX PR 16-FEB-2001; 2001US-0269308P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
XX PI Yamamoto RT, Xu HH;  
XX DR WPI; 2001-611495/70.  
XX DR N-PSDB; AAS52066.  
XX PT New polynucleotides for the identification and development of  
XX PT antibiotics, comprise sequences of antisense nucleic acids.  
XX PS Example 3; SEQ ID NO 5703; 511pp; English.  
XX CC The invention relates to antisense inhibitors of genes essential to  
XX CC prokaryotic cellular proliferation, their use in identifying the genes,  
XX CC their use in the discovery of novel antibiotics, the essential genes,  
XX CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
XX CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
XX CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
XX CC useful for the identification of potential new targets for antibiotic  
XX CC development. The antisense nucleic acids can also be used to identify  
XX CC proteins used in proliferation, to express these proteins, and to obtain  
XX CC antibodies capable of binding to the expressed proteins. The proteins can  
XX CC be used to screen compounds in rational drug discovery programmes. The  
XX CC antisense nucleic acid sequence is also useful to screen for homologous  
XX CC nucleic acids which are required for cell proliferation in a wide variety  
XX CC of organisms. The present sequence represents an essential prokaryotic  
XX CC cellular proliferation protein. Note: The sequence data for this patent  
XX CC did not form part of the printed specification, but was obtained in  
XX CC electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 2025 AA;

Query Match 4.0%; Score 202.5; DB 4; Length 2025;  
Best Local Similarity 18.5%; Pred. No. 0.0046;  
Matches 216; Conservative 174; Mismatches 410; Indels 367; Gaps 55;

QY 16 SVVTHNQEVFLVKEPILKQTSASSISGADYAES--GKSKL-----KINET 61  
Db 413 AAVOHAKDLINQTSNPTLDKQAVSQLTQGVNQAKNLHGDQKLDADKQHAVTDNLQNSL 472  
QY 62 SGPDVDDTVTDLFSDKRTTPEKIKDNLAKGPR-EQELKAVTENTSEKQITSGSQLEQSK 120  
Db 473 NNPRQALLESQINNAATFDE-VAQKLAELAQALDQAMQALRNSIQDQQQTSSGKPF----- 526



Db 1604 NNPOQALQSINNAATRDE-VAQKLABAQAALDQAMQALRNSIQDQQQTSGSKF-----1657  
Qy 121 SLSLNKTPSTNWEIC-----DFTYK-GNTLVGLSKSGVEKLSQT-----DHL-----163  
Db 1658 ---INEDPKQDAYQAQVQAHXDLINOTGNPT--LDKQVEQLTOAVTTAKDNLHGDKL 1712  
Qy 164 -----VLP-----SQAADGTQLIQVASFAPTPDKKTAIAEYTSRAGE-----200  
Db 1713 ARDQQAQVTTNPNLNLHAAQQAALDANAA-----PRTREVAQHVTATLPHAMET 1766  
Qy 201 -NGEISOLD-----VDGKEIINEGEVFNYSYLLKKVTIPTGYKHIGQDAFVD--245  
Db 1767 LKNKYDQVNTDKAQPNYTEASTDKKEAVDQA---LQAAESITDPFNGSNANKAVEQAL 1822  
Qy 246 -----NKMAEVLNPESETISDYAPHL-----ALKQIDLPDLNKAIGEL 286  
Db 1823 TKLOBKVNELNGERNVABAK-AAQAKTIDQ--LAHLNADQIATAKQNTDOATKLOPIABL 1879  
Qy 287 AFFDNQI-----TGKLSLPROLMRLAERAPKSHIK--317  
Db 1880 VDOATQLNQSMQLOQAQVNEHANVETQVDYTDQSDKQNAKYQAIAEAVNLKQNSNQ 1939  
Qy 318 -----TIEFRGSLKVIQBSAQDNDLSQMLPDGLEKTESBAFTGNPGDDHYN 366  
Db 1940 VDOALQNLNKAQALNGDERVALAKTNGK-HDIDQL---NALNNAQDQGFGRIDQSHDL 1995  
Qy 367 NRVVLWTKSGK-----NPSGLATENTYNNPDKSLWQESPE-IDYTKW---407  
Db 1996 NQIQOIVDEAKALNRAMPDLSOEISGNEGRTKGSTNYNADTQVQKVYDEAVDKAQALD 2055  
Qy 408 -----LEEDFTYQKNSVTG--PSNKGLOKVKR-----NKNLETPK 440  
Db 2056 KSTQNLRTABQVILKNDVATAAKKALNGERLNRRKSEALQKLDQLTHLNNAQRLALQ 2115  
Qy 441 QHNGVTITEIGD--NAPRNVDPNQNTLRKY-DLEEVKLPSTIRKIGAFAFOSNNLKS---494  
Db 2116 INNAETLNKASRAINRATKLDNAMGAVQYIDEQHLGVISSTNYNA---DDNLKANYD 2171  
Qy 495 ---FASDDLEIKGAPFNNRIETLELKDKLVTTIGDAAFHINHIYAVLPESVOEIGRS 551  
Db 2172 NAIANAHAHELKVQGNATA--KABABQLKQNIIDAQNAL-----2208  
Qy 552 AFRQGANNLIPMGSKVKTGEMAFSLNRLEHLDSLSEQKQLTEIPVQAFSDNALKEVLLP 611  
Db 2209 -----NGDQNLNKAQKAN-----AFVN-----SLMGLAQ-----2233  
Qy 612 ASLKTIREAFKKNHLKQLEVASALSHTAFNALDNDGDEQP-----DNKVVVYKTHNSYA 667  
Db 2234 -----QQQDLAHNAINNADTVSDVDIVNNQIDLDNDAWETLKLHVDNEIPNAEQTVNYQ 2287  
Qy 668 LADGEHPIVDPKLSSTIVDLBKILKLEGLDYSTLRQTQTQFRDMTTAGKALLSKSNL 727  
Db 2288 NAD-----DNAKTNFDDAKRLA-----NTLNSDNTNVNDINGAIOAVKDAIQN 2331  
Qy 728 RQGEKQKFLQEAQFFLGRVLDLDAKAEKALVTK-----KATNGQOLLERSINKA-V 779  
Db 2332 LNGEOR--LOEAK-----DKALQNVNKLADKLKEIASNATDQKLIK--NKABE 2379  
Qy 780 LAYNNSAIKANVRLKLELDLITGLVBGKGLPQAQTMVQGVYLLKTPLPPEYVIGLVN 839  
Db 2380 LA--NSIIN--NINKATSNQDVSVQQTAGNQAIEQVHANE-----IPKAKIDANK 2425  
Qy 840 YDKSGK-LIYALNMSDITGEQKDAYGNPILNVDEB--NEGYPHALAVATLADYEG-----892  
Db 2426 DVDQKQVQALIDIEGRNPNLTDKEQALKDRINGILQQGHNDINNALTKTEAIEQAKERLAQ 2485  
Qy 893 ---LDIKTLNLSQLSQTSTROVPTAAHYRAGIFQAIQNAAAAEOLLPKPGTHSKSSSS 950  
Db 2486 ALQDINKDLVAKEDAKNKIKAL-----ANAKROQINSNPDLTPQKAKA 2529  
Qy 951 ---ESANSKDRGLQS--NPKT-----NRG 969  
Db 2530 LKEIDBAEKRALQNVENAGTIDQLNRG 2556

## RESULT 44

AAV19935

ID AAV19935 standard; protein; 1087 AA.

XX AC AAV19935;

XX DT 19-JUL-1999 (first entry)

XX DE B. burgdorferi antigenic protein, t742.aa.

XX KW Antigenic protein; vaccine; Lyme disease; infection; detection.

XX OS Borrelia burgdorferi.

XX PN WO9859071-A1.

XX PD 30-DEC-1998.

XX PF 18-JUN-1998; 98WO-US012718.

XX PR 20-JUN-1997; 97US-0050359P.

XX PR 22-JUL-1997; 97US-0053344P.

XX PR 22-JUL-1997; 97US-0053377P.

XX PR 03-SEP-1997; 97US-0057483P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (MEDI-) MEDIMMUNE INC.

XX PI Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX DR WPI; 1999-189980/16.

XX DR N-PSDB; AAX61632.

XX CC New isolated Borrelia burgdorferi nucleic acids - used to develop

XX CC products for the diagnosis, prevention and treatment of diseases caused

XX CC by Borrelia, particularly Lyme disease.

XX PS Claim 12; Page 125; 275pp; English.

XX CC This sequence represents a Borrelia burgdorferi (Bb) protein of the

XX CC invention, which is suitable for use in a vaccine. The Bb polypeptides

XX CC can be used in vaccines for eliciting protective antibodies to members of

XX CC the Borrelia genus, particularly for the use against Lyme disease in

XX CC humans and animals. They can be used for preventing or attenuating an

XX CC infection caused by a member of the Borrelia genus. The products can also

XX CC be used for detection of members of the Borrelia genus

SQ Sequence 1087 AA;

Query Match 3.9%; Score 200; DB 2; Length 1087;

Best Local Similarity 19.6%; Pred. No. 0.0027;

Matches 212; Conservative 164; Mismatches 358; Indels 347; Gaps 54;

Qy 18 VTHNQEVPSLVKEPIILKQTOASSISGADYAESGSKLKLINETSQVDDTVTDLPSDKR 77

Db 156 VNDQKNLFNLEK---LKK-----NLGKSNSENILNDSQKIE-----NDKQ 193

Qy 78 TT---PEKIKONLAKGPREQE-----LKAVTENTESEKQITSGSQLEQSK-----ES 121

Db 194 NTNLSKEKNSENILKTPDNSKYNNNNNTTSLKKISSNSQKSELSPPSQTIIKLYRPS 253

Qy 122 LSLNKTVPSTNSWEICDFTKNTLVG-----LSKSGV-EKLSQTHLVLPQAADGTQ 174

Db 254 YLIKKEL-----YBILDDINTGRVTGLGNRLKELIKGLSNKFKQVKNELIENSKKEASN 308

Qy 175 LI-----QVASFAPTPDKKTAETYSRAGENGEISQDVGKEIINEGEVFNYSY 224

Db 309 LLLTLIKKIDIEPNLINIPKPYK-----EIFOLDKDKK-----POY 346

Qy 225 L-----LKKVTIPTGYKHIGQDAFVD-----NKMAEVLNPESETI 261

Db 347 LEDLSKSVHSIKPIDLENTKSR--QQAIDKLNFLKNPNDAQASKTLAQANKIQHLEDL 404  
QY 262 SDYAFALHAKQIDLPDNLK---AIGELAFF-----DNQITGKLSLPQRLMLAERAF 311  
Db 405 KSKVH---SIKPIDL-ENTKSRQQAIDKLNFLKNPNDAQASKTLAQANKIQHLEDLS 460  
QY 312 KSNHIKTIEF-----RGNLSKVIIGRASFDND-----LSQMLPDGLKIESEAPT 357  
Db 461 KVSIIKPIDLENTKSRQQAIDKLNFLKNPNDAQASKTLAQANKIQHLEDLSKSVHS 518  
QY 358 GNPBGDDHYNRVVLWTGSKNPSGLATENTYVNPDKSLWOESPEIDYTKWLEEDTYQKN 417  
Db 519 IKPID-----LENT-----KSRQQAIDKLN-----EFX-KNN 544  
QY 418 SVTGFSNKGLOKVRKNKLEIPKHQNGVTITEIGNAFRNVDFQNKTLKYDLEEVKLPS 477  
Db 545 PNDQAASKTLAQANKIQHLEDLSK-----VHSIKPIDLENTKSRQQAIDKLN--- 592  
QY 478 TIRKIGAFQSNLKSFEASDOLBEIKEGAFMNNRIETLE-LKOKLVITIGDAAPHINHI 536  
Db 593 -----EFXKNNPNDAQASKTLAQ-----NKIQHLEDLSK-----VHSI 627  
QY 537 YAIVL-----PESQVEIGRSFRQNGANLIFMGSKVKTIGEMAFLSNRLEHLDSEOK 590  
Db 628 KPIDLENTKSRQQAIDKLNFLKNPNDAQASKTLAQ-----ANKIQHLEDLSK 677  
QY 591 QLTETPQAFSDNALKEVLLPASLKTIREAFKNNHLKQLEVASALSHTAFNALDNDGD 650  
Db 678 VHSIKPIDLENTKSRQQAIDKLNFLKNPNDAQASKTLAQ-----ENNGD 725  
QY 651 -----EQFQDNKVVKVT-----HNN-----SYALADGEHFI-----VDPDKLS 683  
Db 726 LLKAENAYEKKIILNTQSDHYKLGIIIRPKLKYHSHIESFDQTIKLPKPKKALHNKI 785  
QY 684 TVIDLEKILKLEGLDYSTLROTTQTFRDMTAGKALLSKNLROGEKQKFLQBAQ--- 740  
Db 786 ALMLNKNKAIESFE-----KAIQIDKNYGTAVYQKGIABEKNGD 826  
QY 741 -----FFLGRVDLQKAIKAEKALVTKKATKNGQLLESIN-----KAVLAYNS 785  
Db 827 MQQAFASFKNAVNLQKPNYALKAGIVSNLGNFKQSEYLNFFNANAKPNEIAIYNLS 886  
QY 786 AIKKANVRLEKLELDLITGLVE-----GKPLAQATWQGVYLLKTP 828  
Db 887 IAKFEN-NKLESLETINKAIDLNPESKSEYLYLKAISNLKENYQNAISLYSVIEKOP- 944  
QY 829 PLPEYIYGLNVDFSGKLIYALD-MSDTIGSQKDAYGN----- 867  
Db 945 ENTSAYINLAKAYERSGKNSQAISTLEKIINKNNKLALNNLGLYKKEKNYQKAIETFEK 1004  
QY 868 PILNVDENEGYHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAYHRAGIFQAIQ 927  
Db 1005 AIIN--SDIEAKYNLATTLEINDNTTRAKDLR-----EYTKLPNNPEALHALGIIENE 1058  
QY 928 N 928  
Db 1059 N 1059

## RESULT 45

AAV19934

XZ AAV19934 standard; protein; 1119 AA.

XX AAV19934;

XX AAV19934;

XX 19-JUL-1999 (first entry)

XX B. burgdorferi antigenic protein, f742.aa.

XX Antigenic protein; vaccine; Lyme disease; infection; detection.

XX Borrelia burgdorferi.

PN WO9859071-A1.  
XX 30-DEC-1998.  
XX 18-JUN-1998; 98WO-US012718.  
XX 20-JUN-1997; 97US-0050359P.  
PR 22-JUL-1997; 97US-0053344P.  
PR 22-JUL-1997; 97US-0053377P.  
PR 03-SEP-1997; 97US-0057483P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA (MEDI-) MEDIMMUNE INC.  
XX Choi GH, Erwin AL, Hanson MS, Lathigra R;  
WPI, 1999-189980/16.  
DR N-P8DB; AAX61631.  
XX New isolated Borrelia burgdorferi nucleic acids - used to develop  
PT products for the diagnosis, prevention and treatment of diseases caused  
PT by Borrelia, particularly Lyme disease.  
XX Claim 12; Page 124-125; 275pp; English.  
XX This sequence represents a Borrelia burgdorferi (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC can be used in vaccines for eliciting protective antibodies to members of  
CC the Borrelia genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the Borrelia genus. The products can also  
CC be used for detection of members of the Borrelia genus  
XX SQ Sequence 1119 AA;

Query Match 3.9%; Score 200; DB 2; Length 1119;  
Best Local Similarity 19.6%; Pred. No. 0.0028;  
Matches 212; Conservative 164; Mismatches 358; Indels 347; Gaps 54;  
QY 18 VTHQEVSLVKEPIKQTOASSISGADYAESKSKLKNETSGPVDVDTTDFSDKR 77  
Db 188 VNDQKLNFLK---LKK-----NLGSKNSNINLDSQKIE-----NDKQ 225  
QY 78 TT---PEKIKNLAKGPREQE-----LKAVENTESEKOITSGSQLEQSK---ES 121  
Db 226 NINLSKEKNSNINLTPNSKYNNNTSLKISSNQSELSFPQTIIKGIYRPS 285  
QY 122 LSLANKTPSTSNWEICDFITKNTLVG-----LSKSGV-EKLSQTDHLVPSQAADGTQ 174  
Db 286 YLIKKE-----YEILDDINTGRVTGKNNRLKELIKGLSNKFKQVKNELIENSKNEASN 340  
QY 175 LI-----QVASFAPTDKTAIAEYTSRAGENCEISQLDVGDKEINEGEVFNYSY 224  
Db 341 LLLTLIKKIDBPFLNINPKDPYK-----EIFQLDKEDK-----PQY 378  
QY 225 L-----LKKVTIPTGYKHIGODAFVD-----NKNIAEVLNPSLEVTI 261  
Db 379 LEDLSKSVHSIKPIDLENTKSR--QQAIDKLNFLKNPNDAQASKTLAQANKIQHLEDL 436  
QY 262 SDYAFALHAKQIDLPDNLK---AIGELAFF-----DNQITGKLSLPQRLMLAERAF 311  
Db 437 KSKVH---SIKPIDL-ENTKSRQQAIDKLNFLKNPNDAQASKTLAQANKIQHLEDLS 492  
QY 312 KSNHIKTIEF-----RGNLSKVIIGRASFDND-----LSQMLPDGLKIESEAPT 357  
Db 493 KVSIIKPIDLENTKSRQQAIDKLNFLKNPNDAQASKTLAQANKIQHLEDLSKSVHS 550  
QY 358 GNPBGDDHYNRVVLWTGSKNPSGLATENTYVNPDKSLWOESPEIDYTKWLEEDTYQKN 417  
Db 551 IKPID-----LENT-----KSRQQAIDKLN-----EFX-KNN 576  
QY 418 SVTGFSNKGLOKVRKNKLEIPKHQNGVTITEIGNAFRNVDFQNKTLKYDLEEVKLPS 477













XX Vancomycin; teicoplanin; MRSA; diagnosis; SA1577; FmtB;  
 KW antibiotic resistance.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 PN WO2003062466-A2.  
 XX  
 PD 31-JUL-2003.  
 XX  
 XX 16-JAN-2003; 2003WO-GB000129.  
 XX  
 XX 16-JAN-2002; 2002GB-0000865.  
 PR  
 XX (UYBR-) UNIV BRISTOL.  
 PA  
 XX Walsh TR, Avison MB, Howe RA;  
 XX  
 XX WPI; 2003-636743/60.  
 DR N-PSDB; ACF05836.  
 XX  
 XX Detecting a strain of *Staphylococcus aureus* having reduced susceptibility  
 PT to vancomycin and/or teicoplanin, comprises detecting the presence of  
 PT mutation(s) leading to loss of function of at least one gene.  
 XX  
 PS Claim 26; Fig 1; 78pp; English.  
 XX  
 XX The present sequence is that of a putative FmtB-like protein encoded by  
 CC open reading frame SA1577 of methicillin resistant *Staphylococcus aureus*  
 CC (MRSA) strain N315. SA1577 has been assigned the provisional gene name  
 CC Mrp on the basis of homology of its sequence to known genes. This gene is  
 CC mutated in vancomycin resistant *Staphylococcus aureus* strain Mu50, the  
 CC mutation leading to truncation of the encoded protein (see ABB62800). On  
 CC the basis of such loss of function mutations, the invention provides  
 CC polypeptides, polynucleotides, probes, primers, antibodies, kits and  
 CC methods for detecting strains of *S. aureus* having reduced susceptibility  
 CC to vancomycin and/or teicoplanin. The methods are rapid, reproducible and  
 CC sensitive, and allow for detection of strains that are fully or  
 CC heterogeneously resistant  
 XX  
 SQ Sequence 2196 AA;

Query Match 3.9%; Score 199; DB 7; Length 2196;  
 Best Local Similarity 19.2%; Pred. No. 0.0084;  
 Matches 217; Conservative 182; Mismatches 387; Indels 342; Gaps 58;

QY 20 HNOEVSFLKEPILKQTOASSISGADYAESSGSKLKINETSGPVDVTDLFSD---- 75  
 DB 1207 YNAKLAENATPDATNDEKNAINTLNDRQQAIESIKQANTNAEVDQAATVAENNIDAV 1266

QY 76 -----KRTTPEKIKDNLAKGREQELKAVTENTESEKQITSGSOLEOSKE-----SLSLN 125  
 DB 1267 QVUVVKKQAARDKITAEVAK--RIEAVKOTPNATDEEKQ-AAVNQINQLKDAQAINQINQN 1323

QY 126 KT---VPSTSNWEI-----CDFITKGNLTVLGSKSGVEKLSQTDHLVPSQAADGTQL 175  
 DB 1324 QTNQDQVDTTQNAVDNVAEAEVVIKPKAIDIEKAVEKQQQIDNSL---DSTNEK- 1379

QY 176 IQVASFAFTPKKTAIAYTSRAGEISQLDVGKEIINEGEVNSVLLKVT--IPT 233  
 DB 1380 -EVASQALAKEKEKALAA-IDQAQTSQVNAATNGVSAIK-----IIQPETKVKEA 1429

QY 234 GYKHIGODAFVNDKNTAEVNLPE-----SLETISDYAFALAL-----KQID 275  
 DB 1430 ARKINQKA---NELRAKINQDKEATAERQVALDKINE--FVNQAMTDTNRTNQVD 1484

QY 276 LPNLKAIGELAFPDNQITGKLSPLQLMLAER-----AFKSNHIKTIETPRGNSLKV 328  
 DB 1485 -DITSQALDSIA-----LVAPEHIVRAARDVAKQYEAKEQIEQAEHATDEEQ 1534

QY 329 IGEASFQDND-----LSQMLPDGLEKIESEFTGNPGDDHNNRVLTWTKSGKNPSGL 382  
 DB 1535 VALNQLANNEKLALQINQAVTNNVKKRVETNGIATLKG---VQPHIVI-----KPEAQ 1586

QY 383 ATENTYVNPDKSLWOESP-----EIDYTKWL-----BEDFTYQKNSVTGFSN--- 424  
 DB 1587 AIKATAENQVESI-KDTPHATVDELDEANQLISDILKQAQQEIENTNQDAAVTDVNRQTI 1645

QY 425 KGLQVK-----RNKN-----LEIPKOHNGVITTEIGD--NAFRNVD 459  
 DB 1646 KAIEQIKPKVRRKRAALDSIENNNKQLDAIRNTLDTTQDERDVAIDTLNKIVNTIKNDI 1705

QY 460 FONKTLRKVDLEE-----VKL-----PSTIRKIGAFAFQSNNLKSFEASD-DLEEI 504  
 DB 1706 AQNKTAENVDRTERETGNDNRIKVLKPVQVKKPAAROSVGKAEQAQNAL--IDQSDIESTE 1763

QY 505 KEGA-----FMNRIETLELKLVTIGDAAFINHIVAILVP-----ESVQETG 549  
 DB 1764 RLAAKHLVEQALNQALDQINHADKTAQVNNQDSIDAQNIISKIKPATTVKATALQIQNTA 1823

QY 550 RSAFRONGANN-----LIFGSKVK---TLGEMAF-----SNRLEHLD 585  
 DB 1824 TNKINLIKANNEATDDEQNIAIAQVEKELIKAKQOIASAVTNADVAYLLHDEKNEIRETE 1883

QY 586 -----LSEQKOLTEIPQAQPSDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSHI 639  
 DB 1884 PVINKASAREQLTTL-----FND---KKQAIEANTQATVEE--RNSILAQLQ---NIYDT 1931

QY 640 AFNALDDNDGDQFDN--KVVVTKHNSVALADGHEFIVDPDKLSSTIVDLKILKILIEG 697  
 DB 1932 AIGQIDQDRSNAQVNDKTASLNQTIHDL-----DVHPKPPDAEKTINDDLARVTALVO- 1985

QY 698 LDYSTLRQTTQTPFRDMMTAGKALLSKNLRQGEKQFLQEAQOFFLGRVLDLKAIAKAEK 757  
 DB 1986 -----NYRKVSNRNK-----ADALK 2000

QY 758 ALVTKKATKNGOLLERSIN---KAVLAYNNSAIKANVVKLEKELDLLTGLVEGKGPLAQ 814  
 DB 2001 AITALKLQMOEELKTARTNADVDVLRKFNVALSIEAVITEKNSLLR--IDN---IAQ 2055

QY 815 ATMVQGVLLKTPPLPEYVIGLVYFDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDE 874  
 DB 2056 QT-----YAKPKAIATPEQLAKVKVLIQO-----YVAD-----GNRM--IDE 2090

QY 875 DNEGHALAVATLADYEGDLIKTILNSKLSQTSIRQVPTAAHYHRAGIFQAIQNAAAEAE 934  
 DB 2091 D-----ATLN-----DIKQHTQFIVDEILAIAK-LPAEA-----TKVSPKEI 2125

QY 935 QLLPKPGCTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG 982  
 DB 2126 QPAPKVCTPIKKBETHESRKVEK-----LPTGTSEG 2157

RESULT 52  
 ABB63519  
 ID ABB63519 standard; protein; 2346 AA.  
 XX  
 AC ABB63519;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 17349.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US009231.  
 PF  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR  
 PR 11-JUL-2000; 2000US-00614150.





XX Disclosure; SEQ ID NO 1097; 17pp + Sequence Listing; English.  
XX The invention relates to multiprotein complexes from eukaryotes. Proteins  
CC of the invention and DNA sequences encoding them are given in records  
CC ABR52568-ABR53903 and ACC6010-ACC61944 respectively. The complexes are  
CC obtainable by using a protein as a bait and isolating the set of proteins  
CC which is attached thereto from cells. Such protein complexes may comprise  
CC up to 30 distinct proteins. Protein complexes of the invention are useful  
CC for diagnosing a disease or disorder, or as a target for an active agent  
CC of a pharmaceutical, preferably a drug target in the treatment or  
CC prevention of a disease or disorder. Note: the sequence data for this  
CC patent is not represented in the printed specification, but is based on  
CC sequence information supplied by the European Patent Office. The complete  
CC document is available on CD-ROM  
XX  
SQ Sequence 1790 AA;

Query Match 3.9%; Score 197.5; DB 6; Length 1790;  
Best Local Similarity 21.0%; Pred. No. 0.0078;  
Matches 248; Conservative 174; Mismatches 425; Indels 333; Gaps 59;  
QY 25 FSLVKEPILKQQA--SSISGADYAESGSKLKNITSGPVDVTD----- 71  
DB 434 FDLQDFLLKQVLQNNSTNVGDNAKNGSKSDKSDKDTGDKGTGYEGFKA 493  
QY 72 LF-----SDKRTTPEKI-----KDLAKGPREQLKAVTENTSEKQITSG 112  
DB 494 LFEVLLNYDAELNLPFKLFPTTDFMPFFQDQ-----KYSELREITRN-----VTTG 543  
QY 113 SOLQESKESLSLNTVP-----STS-----NWEICDFITKGNLTVLGSKS 152  
DB 544 NDLE-DEEPLKAIQITISELLTSLTAADIRIPISYLTFLIYWLFGDFKATNDFL--SDKS 600  
QY 153 GYEKLSQTHVLVPSQAD-----GTQLIOVA--SPAFTPDKTAIAEY--TSRAGEN- 201  
DB 601 VTKSLLSFSYQI---QDEVTIKCLVTLMLGVAYEFSSKESPPRKEPFETIKLGDN 657  
QY 202 --DETSLQDVD--GKEIINEGEV-----PNSYLLKKVT-----IPTGYKH 237  
DB 658 YASRIKQPKDSYFYSKVDNEDSILTPDELDEGLPKVYFTYFIQFNENIVRIITALSH 717  
QY 238 IQQDAFVDNKNIAEVLNPE-----SLETISDYAFALHAKLDLPDLNKAIGEL 286  
DB 718 DPDEEPINKISPEEVEKLQRCQTKLGBITSLQTETESTHENLTKLIAITNEHKELDEK 777  
QY 287 AFQDN-----QITGKLSLPRQLMRLAERAPKSHNIKTIFRGNLSKV 328  
DB 778 YQILNSSHSLKENFSILETELKNVRDSIDEMTQLRDVLETKDKENQTALEYKSTIHK- 836  
QY 329 IGESAFQDNDLQMLPDGLEKIESAFTGNPGDDHNNRVLVLTGSKGNPSGLATENTY 388  
DB 837 -----QESIK--TLEKLETLISQK--KKAEDGIN-----KMGKDLFALREMOQA 878  
QY 389 VNPD--KSLMQESPEIDY-----TKWLEEDFTYQKNSVTGFSNGKLOKVK--RNKNLEIPKQ 441  
DB 879 VEENCNQLKEXDKNVNVHOKETSKLKEDIAAKITEIKAI--NENLEEMKIQCNNLSKEKE 937  
QY 442 HNGVITEIGD-----NAFRNVDFONTKRLKYDEEVLKLPSTR- 480  
DB 938 HISKELVEYKSRFQSHNDLVAKLTKELKSLANNYKDMQAEENSLIK-AVEESKNESLIQ 996  
QY 481 -----KIGAFQSNLKK-----SFEAS-----DDLEEIKEGAPMNNRIETLE----- 518  
DB 997 SNLQNKIDMSQENKFNQTERGSIEKNIEQLKKTISDLQTKTEETISKSDSKDEYSGOI 1056  
QY 519 --LKDCLVITGDA--APHINHIYAVLPESVQIGRSAPFQNGANNI--IPMGSKVKTLLGEM 574  
DB 1057 SLLEKLELTATTANDENVNKISLTKTRELEAEALAAVK-NLKNLETKLTSEKALKEV 1115  
QY 575 AFLSNRLEHLDLSEKQLTEIPVQAFSDNALKEVL-----LPASLKTIREFAFKKNHLK 628  
DB 1116 KENEELHLEBEKIQLEKEATETKQQLNSLRANLESKEHEDLAAQLKKYEEQIANKERQY 1175

QY 629 QLEVASALSHIAFNALDNDGDGEQFDNKVWVVKTHNSVALADGEHFIV----- 676  
DB 1176 NEEISQ-----LNDEITSTQENESIKKNDE-----LEGEVKAMKSTSEEQSNLKK 1222  
QY 677 -DPDKLSSTIVDLEK-----ILKLJEGLDYSTL-----RQTQTQPRDMTT 716  
DB 1223 SRIDALNLQIKELKKKNETNEASLIESIKSVESETVKIKELQDECNFKEVSELEDKLK 1282  
QY 717 AKKALLSKSNLRQGEKQFLQ-----EAQFFGRV-DLDKAIKAEKALVTKAT--- 765  
DB 1283 ASEDKNSKYLELQKSESEKIKBELDAKTTELKIQLEKITNLSKAKESSELSRLKKTSS 1342  
QY 766 --KNG-QULERSINKAVLAYNNSAIKKANVRLEKELDLITGLVSGKGPQAQATVVG 822  
DB 1343 ERKNAEEQLEKLNKNE--IQIKNQA-----FEKERKLN--EGSSTITQ----- 1381  
QY 823 LLKTPPLPPEYVIGLVNVPDSSGKL-----IYALDMSDTIGEGQKDAYGNPILNVDE 877  
DB 1382 -----EYSEKINTLEDELIRIQENENELKAKEIDNTRSELEKVSLSNDELBEKQN- 1431  
QY 878 GVHALAVATLADYEGLDIKTILNSKLSOLTSTI-----ROVPTAAVHRAGIPOAIONAA 932  
DB 1432 -----TIKSLQD--ELISYKDKITRNDEKLLSTIERDNKRDLSEKQ-----LRA 1481  
QY 933 AEQLLPKPGTHSEKSSSESSESA-----NSKD--RGLQSNPKTN 967  
DB 1482 VEEGLKK-----LEBESSKEKAELEKSKEMMKLESTIESN 1517  
RESULT 54  
ADK63040  
ID ADK63040 standard; protein; 1790 AA.  
XX  
AC ADK63040;  
XX  
XX 06-MAY-2004 (first entry)  
XX Disease treating protein complex-derived protein #646.  
XX protein complex; drug target; diagnosis.  
XX Unidentified.  
XX EP1338608-A2.  
XX 27-AUG-2003.  
XX 20-DEC-2002; 2002EP-00102902.  
XX 20-DEC-2001; 2001EP-00130253.  
XX (CELL-) CELLZOME AG.  
XX Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;  
PI Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;  
PI Michon A, Leutwein C, Rick J;  
XX WPI; 2003-638460/61.  
DR N-PSDB; ADK63041.  
XX  
PT New proteins and protein complexes from eukaryotes, useful as targets in  
PT drug screening, or in diagnosing or screening for the presence of a  
PT disease or disorder, or a predisposition for developing a disease or  
PT disorder in a subject.  
XX Disclosure; SEQ ID NO 1291; 13pp; English.  
XX  
XX The invention relates to novel protein complexes comprising a first and a  
CC second protein, or its derivative, fragment, homologue or variant. The  
CC proteins are selected from given protein complexes, which are not defined  
CC in the specification. The variants are encoded by nucleic acids that  
CC hybridize to the nucleic acids encoding the proteins under low stringency



CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition, improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polypeptide used in the  
 CC scope of the invention. Note: The sequence data for this patent did not  
 CC form part of the printed specification but was obtained in electronic  
 CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).

XX Sequence 1790 AA;

Query Match 3.9%; Score 197.5; DB 8; Length 1790;

Best Local Similarity 21.0%; Pred. No. 0.0078;

Matches 248; Conservative 174; Mismatches 425; Indels 333; Gaps 59;

QY 25 FSLVKEPILKOTQA---SSSISGADYAESGSKLKINETSQFVDDTVTD-----71  
 DB 434 PDLQRDFLLKQVLCNNSTNNVGDNAKENGSGNSKSDSKDGTGDKGTVEYGSFKAN 493  
 QY 72 LF-----SDKRTTPEKI-----KONLAKGPREQELKAVTENSEKQITSG 112  
 DB 494 LFEVLNLDYDAELNLPFKLFFTTFIMFFFOODH-----KYSELREITRN-----VTTG 543  
 QY 113 SQLEQSKESLSLNTKVP---STS-----NWEICDFITKGNLVLGSKS 152  
 DB 544 NDLE--DEPLKAIQIISILLTSLTAADIRIPISYLTFLIYWLFGDFKATNDFL---SKKS 600  
 QY 153 GVEKLSQTDHLVPSQAAD-----GTOLIOVA---SFAFTPDKKAIAEY--TSRAGEN- 201  
 DB 601 VIKSLSPSYQI---QDEVTIKCLVTLMLGVAYEPSSKESPPRKEYFEFTITKLGRON 657  
 QY 202 --GEISQLDVD---GKEIINEGEV-----FNSYLLKKVT-----IPTGVKH 237  
 DB 658 YASRIKQPKDSYFSKVDNEDSILTPDELDTGLPKVYFSTYFIQLFNEINIRIRIALSH 717  
 QY 238 IQQDAFVNKNIAEYNLPE-----SLETISDYAFAHALKQIDLPDLNKAIGEL 286  
 DB 718 DPDEEPINKISPEEVKQROCTKUGETSITQETESTHENLTKLIALTNHEKDELDEK 777  
 QY 287 AFFDN-----QITGKLSLPRQLMLAERAFKSNHINKTIEFRGNSLKV 328  
 DB 778 YQILNSSHSLKENFSILETELKQNVDSLDENTQLRDVLETKDKENQTALLBYKSTIHK- 836  
 QY 329 IGESAFQNDLSQLMPLDGLKEIESEAFNGPDGHHNNRVVLTWKSGNPSGLATENTY 388  
 DB 837 -----QEDSIK--TLEKLEITLSQK-----KKAEDGIN-----KMGKDLFALSREMQA 878  
 QY 389 VNPD--KSLWQESPEIDY-----TKWLEEDFTYQKNSVTGFSNKGKQVK--RNKNLEIPKQ 441  
 DB 879 VEENCKNLQKDKSNVNHQKTKSLKEDIAAKITEIKAI--NENLEEMKIQCNNLSKEKE 937  
 QY 442 HNGVITTEIGD-----NAFRNVDPQNTKLRLKYLEEVKLPSTIR- 480  
 DB 938 HLSKELVEYKRFQSHDNLVAKLTEKLKSLANNYKDMQENESLIK-AVEESKNESLQL 996  
 QY 481 -----KIGAFAPQSNLKK-----SFPAS-----DLEEEIKEGAFMNNRIETLE-----518  
 DB 997 SNLQNKIDMSQEKENFQIERGSIIEKNIEQLKTTISDLQETKEEIIISKSDSKDBYESQI 1056

QY 519 --LKDKLVITGDA-AFHINHIYAIIVLPESVQIGRSAPFRQGANNL-IPMGSKVKTIGSM 574  
 DB 1057 SILKEKLETATTANDENVNKISELTKTREELAEALAAVK-NLKNELTETKLTSEKALKKEV 1115  
 QY 575 AFLSNRLHDLSEQKQITETIPVQAFSDNALKEVL-----LPASLKTIREFAFKXKHUK 628  
 DB 1116 KENEHLKEEKTQLEKEATETKQQLNSLRANLESLEKEHEDLAAQLKKYEQIANKEROY 1175  
 QY 629 QLEVASALSHAFNALDDNDGDEQDPNKVVVKTHNSVALADGEHFIV-----676  
 DB 1176 NEEISQ-----LNDEITSTQOENESIKKONDE-----LEGEVKAMKSTSEESQNLKK 1222  
 QY 677 -DPDKLSSTIVDLK-----ILKLEGLDYSTL-----RQTTOTQFRDMTT 716  
 DB 1223 SEIDALNLIQELKKKNETNEASLLESKSVESETVKIKELQDECNFKKEVESELEDKUK 1282  
 QY 717 AGKALLSKSNLRQGEKQKFLQ-----BAOFFLGRV-DLDKAIAKAEKALVTKKAT---765  
 DB 1283 ASEDKNKYLELQKESEKIKEELDAKTTTELKIQLEKITNLKAKEKESELSRLKKTSE 1342  
 QY 766 --KNG-QLLERSINKAVLAYNNSAIKANVKLEKELDLTLGLVEKGKPLAQATWVGQY 822  
 DB 1343 ERKNAEQLEKLNK--IQKNQA-----FEKERKLLN---EGSSTITO-----1381  
 QY 823 LLKTPPLPPEYVIGLVNVPYFDKSGKL-----IYALDMSDTIGEGQKDAYGNPILNVDEDN 877  
 DB 1382 -----EYSEKINTLEDELIRLONENELKAKEIDNTRSELEKVSLSNDELSEKON- 1431  
 QY 878 GYHALAVATLADYEGLDIKTILNSKLSQTSI-----ROVPTAAYHRAGIFOAIQNAAB 932  
 DB 1432 -----TIKSLQD-ETLSYKDKITRNDKLSLSTERNKRDLESKEQ-----LRAAQESKAK 1481  
 QY 933 AEOLLPKPGTHSEKSSSESSESA---NSKD--RGLQSNPKTN 967  
 DB 1482 VEEGLK---LEESSEKKEALEKSKEMMKKLESTIESN 1517  
 RESULT 56  
 AAU34139  
 ID AAU34139 standard; protein; 2368 AA.  
 XX  
 AC AAU34139;  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Staphylococcus aureus cellular proliferation protein #415.  
 KW Antisense; prokaryotic cellular proliferation protein; antibiotic;  
 KW antibacterial; drug design.  
 OS Staphylococcus aureus.  
 XX  
 FN WO200170955-A2.  
 XX 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-0009180.  
 XX  
 PR 21-MAR-2000; 2000US-0191078P.  
 PR 23-MAY-2000; 2000US-0206848P.  
 PR 26-MAY-2000; 2000US-0207727P.  
 PR 23-OCT-2000; 2000US-0242578P.  
 PR 27-NOV-2000; 2000US-0253625P.  
 PR 22-DEC-2000; 2000US-0257931P.  
 PR 16-FEB-2001; 2001US-0269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlson KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI; 2001-611495/70.

DR N-PSDB; AAS51998.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids.

XX

PS Example 3; SEQ ID NO 5635; 511pp; English.

XX

CC The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the genes,

CC their use in the discovery of novel antibiotics, the essential genes

CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*

CC coli, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,

CC *Pseudomonas aeruginosa*, and *Enterococcus faecalis*. The invention is also

CC useful for the identification of potential new targets for antibiotic

CC development. The antisense nucleic acids can also be used to identify

CC proteins used in proliferation, to express these proteins, and to obtain

CC antibodies capable of binding to the expressed proteins. The proteins can

CC be used to screen compounds in rational drug discovery programmes. The

CC antisense nucleic acid sequence is also useful to screen for homologous

CC nucleic acids which are required for cell proliferation in a wide variety

CC of organisms. The present sequence represents an essential prokaryotic

CC cellular proliferation protein. Note: The sequence data for this patent

CC did not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 2368 AA;

Query Match 3.9%; Score 197.5; DB 4; Length 2368;

Best Local Similarity 19.8%; Pred. No. 0.011;

Matches 242; Conservative 165; Mismatches 459; Indels 359; Gaps 59;

QY 19 THNQVSLVEPIKQTOQASSISGADYASSG-----KSKUKINETSGPVDVDTVD 71

DB 850 TQNTATVSNATNEAEADAA-----VEAAQKQGLHDIQVVKSKQEVADTKSVLKDINA 903

QY 72 LPSDKRTTPEKIKO-NLAKGPREQELKAVTENTSEKQITSGSQLESKESLNTKTPVS 130

DB 904 IQTQAKVPAADTEVENATYRKQBIQNSNASTTEEKQ-AAYTELDITKKQBARNTLDAAN 962

QY 131 TSNMEICDPIT-KGNTLVGLS-----KSGVEKLS----- 158

DB 963 TNS-----DVTTAKONGIAAIQVQAATTKSDAKAEIAQKASERKTAIEAMNDSTTEEQ 1018

QY 159 ----QTDHLVLPQSA-----ADQTLIQVASF-AFTPD-----KKTAI-----A 192

DB 1019 AAKDKVDQAVVTANADIDNAAANTDVDNAKTNEATIAAITPDANVVKPTAKQAIADKVOA 1078

QY 193 EYTSRAGENGISQLVDGKEINEGEVFNLSVLLKKVTIPTGYKHIGQDAFVD---NKNV 249

DB 1079 QETAIDANGATTEEKAQKQVQ-----TEKTTADTAIDGAHTNAEVEAAKNABE 1129

QY 250 AEVNLPSLETISDYAFAPHLALQIDLPDLKAIGEL-----AFFDQITGKLSLP 300

DB 1130 AKIEAIPQATTKONAKQAIATKANERKTAIAQTQDITABEIAAANAVDNAV----- 1183

QY 301 RQLMLAEAPKSNHIKTIIEFRGSLKVIIGRASFOQNDLSQMLPDGLEKTESEFTGN- 359

DB 1184 -QANNNIEAANSQNDV-----DOAKTTGEASIDQ-----VTPVNVKKAATVTDKNN 1229

QY 360 ---PGDDHNNRVNVLWTKSGKN-----PSGLATENTVNPDKSLWQESPEIDVTKWLEE 410

DB 1230 ITAATDD---NGVDTAKDAGKNSIQSTQFATVAKSNKNDVDQAVTTQQAIDNTGTAT- 1285

QY 411 DFTYQKNSVTGFSNKGLOKVRKNKLEIPKQHGVTITEIGDNAPRNV-----DFQNKTL 465

DB 1286 ---TEEKNAKDL-----VLKAKEKAQDILNQAQTTNDVTQIKDQAVADVQGITADTTIKDV 1339

QY 466 RKYDI-----EVLKPGSTI-----RKIGAFAPQSN-NLKSFEASDDLEIEKEG 507

DB 1340 AKDELATKAREQKALIAQTADATTEBEQANQVDAELTQGNQNTIENASQSIDDVNTAKDN 1399

QY 508 AFM-----NNRIETL-ELKDKLVITIGDAAFHINHIYAIVLPSVQIEG--RS 551

DB 1400 AIQAIDPIQASTDVKTNARAELELTEMQKTEILNNNTTNE-----EKGNDIGPVRA 1452

QY 552 AFRQGANNLIFMGSKVKTKLGEAMFLSNRLEHLDLSQKQLTEIPVQAFSDNALKEVLLP 611

DB 1453 AY-BEGLNNI-----NAATTTG-----DVTTAKDTAVQKVQQLHANPVKK---P 1492

QY 612 ASLTKTIREAP--KKNHLKQL-----EVASALSHIAFNALDDNDGDEQFPNK 656

DB 1493 AG-KTALDQAAADKKTOIEQTPNASQOEINDAQOEVDTELNQAKTN-IDQSSTDEYVDNA 1550

QY 657 V-----VVKT--HNSVALADGEHF-----IVDPD-----KLSSTVDL-EKILKLEIG 697

DB 1551 VKEGAKINAVKTPSEYKKAALAKIEAAYNAKVTEADNSASTSEIAEAKQKLAELKQT 1610

QY 698 LDYSLRQLTQTQPRDMMTAGKALLSKNSLRQGEKQKFLQEAQFFL-----GRVDLDKAIA 753

DB 1611 ADQNVNQTAKSKDIEVQIHNDLDNIINDYTIPTGKESATTDLYAYADQKKNNISADTNAT 1670

QY 754 KAEKALVTKATKNGQLLERSINKAV-----LAYNSAI-----KCAN----- 791

DB 1671 QDERQQAIKQVDQNVQTALESINNGVDVDDALTOGKAIDAIOVDATVVKPKANQVID 1730

QY 792 ----VKRLEKELELLTG-----LVGKGPLAQATMVQGVYLLKTLPLPLEY 833

DB 1731 AKABETKESIDQSLTAEKTEALAMIKQITDQAKQGITDATTAEVEKAKAQ----- 1784

QY 834 YIGL-----NVYFDKSGK-----LIYALDM-----SDTIGRGQKDAYGNPILNVDED 875

DB 1785 --GLEAFDNIQIDSTEKQKAELEETALDQIEAGVNVVDADATTE-EKEAFTNAL----- 1835

QY 876 NEGTHALAVATLADY-BGLDIKTLNSKLSQTSIRQVPTAAYHRAGI-----FOAI 926

DB 1836 -EDLSKATEDISDQTTNAEIAIVKNSALEQKACQINPVVVKNALEAIREVVNKKQIEII 1894

QY 927 QNAAABAE-----QLLPKPGTHSEKS-----SSSESAN 954

DB 1895 KNADADASAKEIARTDLGRYDFRPAKLDKDTQTNTIEVAELQNVTIPIAEIIVPQNDPNAN 1954

QY 955 SKDRGLQSNPKNRRHSAILPRTG 979

DB 1955 DTNSGSDNDNATANSNANATPENTG 1979

RESULT 57

AAU36796

ID AAU36796 standard; protein; 2368 AA.

XX

AC AAU36796;

XX

XX 14-FEB-2002 (first entry)

DT

XX

DE Staphylococcus aureus cellular proliferation protein #96.

XX

KW Antisense; prokaryotic cellular proliferation protein; antibiotic;

KW antibacterial; drug design.

XX

OS Staphylococcus aureus.

XX

PN WO200170955-A2.

XX

PD 27-SEP-2001.

XX

XX 21-MAR-2001; 2001WO-US009180.

PF

XX 21-MAR-2000; 2000US-0191078P.

PR 23-MAY-2000; 2000US-0206848P.

PR 26-MAY-2000; 2000US-0207727P.

PR 23-OCT-2000; 2000US-0242578P.

PR 27-NOV-2000; 2000US-0253625P.

PR 22-DEC-2000; 2000US-0257931P.

PR 16-FEB-2001; 2001US-0269308P.

XX

(ELIT-) ELITRA PHARM INC.

PA Haselbeck R.; Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX

XX WPI; 2001-611495/70.  
 DR N-PSDB; AAS54655.  
 XX

PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids.  
 XX

PS Example 3; SEQ ID NO 12389; 511pp; English.  
 XX

CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the genes,  
 CC their use in the discovery of novel antibiotics, the essential genes  
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
 CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence represents an essential prokaryotic  
 CC cellular proliferation protein. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX

Sequence 2368 AA;

Query Match 3.9%; Score 197.5; DB 4; Length 2368;

Best Local Similarity 19.8%; Pred. No. 0.011;

Matches 242; Conservative 165; Mismatches 459; Indels 359; Gaps 59;

QY 19 THNQEVSLVLPKILKQTOASSISGADYAESG-----KSKLKINETSQVDDTVTD 71  
 DB TQNTVSNATNEVEAEADAA-----VEAQQLGHLDIQVKSQGEVADTKSKVLDDKINA 903  
 QY 72 LFSDRRTTPEKID-NLAKGPREQELKAVTENTSEKQITSGSQLESKESLSNKTVPES 130  
 DB IQTOAKVKAADTEVENAYNRKQEIQNSNASTEKQ-AAYTELDTKKQEARTNLDRAN 962  
 QY 131 TSNWEICDPIT-KGNVLVGLS-----KSGVEKLS----- 158  
 DB TNS-----DVTAKDNGIAINOQVQAATTKKSDAKAEIAQKASERKTAIEAMNDSTTEQ 1018  
 QY 159 -----QTDHLVLPSEA-----ADGTQLIQVASF-AFTPD-----KKTAI-----A 192  
 DB 1019 AAKDKVDQAVVANADIDNAAANTVDVNNAKTNEATIAAITPDANVKTAKQAIADKQVA 1078  
 QY 193 EYTSRAGENGEISQLDVDGKEIINEGEVFNYSLLKKVITPTGYKHIGQDAFVD---NKN 249  
 DB 1079 QETAIIDANNATTEKAAKQVQ-----TEKTTADTAIDGAHTNAVEAAKNAEI 1129  
 QY 250 AEVNLPESETSDVAFALHALKQIDLPLNLKAIGEL-----AFPDNOITGKLSLP 300  
 DB 1130 AKIEAIQPAATTTKDNAKQAIATKANERKTAIAQTQDITAEETAAANANVNDNAV 1183  
 QY 301 ROLMLAEAFKSNHIKTIETFRGNSLKVIGESFQDNDLSQLMPLGLEKIESEAF 359  
 DB 1184 -QANNNIEAANSQNDV-----DQAKTGEASIDQ-----VTPTVKKKATAVTDKKN 1229  
 QY 360 ---PGDDHYNNRWLWTKSGKN-----PSGLATENTYVNPDKSLWQESPEIDYTKWLEE 410  
 DB 1230 ITAATD---NGVDYAKDAGKNSIQSTQATAVKSNKNDVDQAVTTQQAIDNTGAT- 1285  
 QY 411 DFTYQKNSVTGFSNKGLOKVKRNKLEIPKQHNGVYITIEIGDNAPRV-----DFQNKTL 465  
 DB 1286 ---TEEKNAAKDL-----VLKAKEKAVQDILNAQTNDVTQIKQAVADVQGITADTTIKDV 1339

RESULT 58

AAB18161

ID AAB18161 standard; protein; 2441 AA.

XX

AC AAB18161;

XX

DT 07-NOV-2000 (first entry)

XX

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:18.

XX

KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;

XX

KW antimalarial; malaria; protozoacide; infection; insecticide.

XX

OS Plasmodium falciparum.

XX

PN WO200025728-A2.

XX

PD 11-MAY-2000.

XX

PF 05-NOV-1999; 98WO-US026796.

XX

PR 05-NOV-1998; 98US-0107131P.

XX





FT Region 560. .581  
/label= tandem repeat  
/note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 582. .603  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 604. .625  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 626. .647  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 648. .669  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 670. .691  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 692. .713  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 714. .735  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 736. .757  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 758. .779  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 780. .801  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 802. .823  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 824. .845  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 846. .867  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 868. .889  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 890. .911  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 912. .933  
/label= tandem repeat  
FT /note= "one of 21 complete segments of homology of 22  
FT amino acid length"  
FT 934. .946  
/label= partial\_tandem\_repeat  
FT /label= tandem repeat  
FT /note= "one of 11 complete segments of homology of 19  
FT amino acid length"

FT Region 968. .986  
/label= tandem repeat  
FT /note= "one of 11 complete segments of homology of 19  
FT amino acid length"  
FT 987. .1005  
/label= tandem repeat  
FT /note= "one of 11 complete segments of homology of 19  
FT amino acid length"  
FT 1006. .1024  
/label= tandem repeat  
FT /note= "one of 11 complete segments of homology of 19  
FT amino acid length"  
FT 1025. .1043  
/label= tandem repeat  
FT /note= "one of 11 complete segments of homology of 19  
FT amino acid length"  
FT 1044. .1062  
/label= tandem repeat  
FT /note= "one of 11 complete segments of homology of 19  
FT amino acid length"  
FT 1063. .1081  
/label= tandem repeat  
FT /note= "one of 11 complete segments of homology of 19  
FT amino acid length"  
FT 1082. .1100  
/label= tandem repeat  
FT /note= "one of 11 complete segments of homology of 19  
FT amino acid length"  
FT 1101. .1119  
/label= tandem repeat  
FT /note= "one of 11 complete segments of homology of 19  
FT amino acid length"  
FT 1120. .1138  
/label= tandem repeat  
FT /note= "one of 11 complete segments of homology of 19  
FT amino acid length"  
FT 1139. .1157  
/label= tandem repeat  
FT /note= "one of 11 complete segments of homology of 19  
FT amino acid length"  
FT 1158. .1173  
/label= partial\_tandem\_repeat  
FT /note= "partial segment of homology"  
FT 1179. .1193  
/label= tandem repeat  
FT /note= "one of 4 complete segments of homology of 15  
FT amino acid length"  
FT 1194. .1208  
/label= tandem repeat  
FT /note= "one of 4 complete segments of homology of 15  
FT amino acid length"  
FT 1209. .1223  
/label= tandem repeat  
FT /note= "one of 4 complete segments of homology of 15  
FT amino acid length"  
FT 1224. .1238  
/label= tandem repeat  
FT /note= "one of 4 complete segments of homology of 15  
FT amino acid length"  
FT 1248. .1260  
/label= tandem repeat  
FT /note= "one of 27 complete segments of homology of 13  
FT amino acid length"  
FT 1261. .1273  
/label= tandem repeat  
FT /note= "one of 27 complete segments of homology of 13  
FT amino acid length"  
FT 1274. .1286  
/label= tandem repeat  
FT /note= "one of 27 complete segments of homology of 13  
FT amino acid length"  
FT 1287. .1299





Query Match		3.9%; Score 196.5; DB 7; Length 1875;
Best Local Similarity		20.8%; Pred. No. 0.0095;
Matches		242; Conservative 164; Mismatches 378; Indels 379; Gaps 59;
Qy	34	KQTQASSSI--SGADYAESGSKLKINET---SGFVDDTDTDL-----PSDKRT 78
Db	199	RKTQELTLQSNNDWLEKELRSK---NEQYLSYRQKTDKVILDIRNELNRLRNDPQMERT 255
Qy	79	TPEKIK-----DNLAKGPRQ--ELKAVTENTSEKOITSG-----SOLEOSKR 120
Db	256	NNDLVKQKNNELSKLQELKEIKGLSDLSNSEQEFAEMSLKQRLVDLLESQNAVKE 315
Qy	121	SLS-----LN--KTPVSTNWEICDPIITKGNLTGLVLSKGVKLSQTDHLVLPQA----- 169
Db	316	ELNSIRELNTAKVTADDSKKQ---TPNEDLLKELQTKELKLAQCEKCELRLLSSITDEA 371
Qy	170	-ADGTQLTQVASFAPTPDKTAIABYTSRAGENGESQLDDVGKE---IIN----- 216
Db	372	DEDNENLSAKSSDPIFLKQLIKERRTKEHLQNIETFIVELEHKVPIINSFKERTDML 431
Qy	217	EGEVNFSYLLKKVTIPTGYKHIGQDAFVDNKNIAFVNLPESLETISDYAFALHALKQIDL 276
Db	432	ENELNNAALL-----LEHTSNE---KNAKVKELNAKQ-----KLVEC 466
Qy	277	PDNLKAIGELAFFDQITGKLSLPQMLAERAFKSNHIK-----TIEP----- 321
Db	467	ENDLQTLTK-----QRDLCKQIQYLIITNSVNSDNGPRLKEBIEIQIOMQEDD 517
Qy	322	---RGNLSKVIGASFQNDLSQLM-----LPDGLE-----KIESE 354
Db	518	STITESDQKVVERTELVFKNIIQIQKNAELLKVVRNLADKLESKEKSKQSLQKIESE 577
Qy	355	AFTNGPGDDHNNRVLMWTKSGQNPGLATENTYVNPDKSLWQSPESPIDTYKWLEEDPT 414
Db	578	-TVNEAKE-----AIITLKSEK-----MDLESRIEELQLEELKTS 613
Qy	415	QKNSVTGFSNKGLOKVXENK-NLEIPKQHGVITIEIGNAFRNVDFONKTLRKYVDLEEV 473
Db	614	VPNEDASVNTIKQLTETKDDLESQVQDLOTRISQITRESTENSLINKIEIQ---DLYDS 671
Qy	474	KLPTIRKIG-----AFAPQSNLKSFEASDDLEIKEGAFMNNRI----- 514
Db	672	KSDISI--KLGEKSSRIILAEERFKLLSNTLDLTRAEND-QLRKRPDYQLNTILKQDSKTH 729
Qy	515	-----ETLEKOKLVITIGDAAFI--NHIYAVLPESV-QEIGRSAPFRONGANNLIFMGSK 567
Db	730	ETLNEYVSCSKLSIVETELLNLKBEQKLRVHLEKNLQELNKLSPK---DSLRIWVTQ 786
Qy	568	VKTIGEMAFNLRLHLDLSQKQLTETPVQAFSDNALKEVLLPASLAKTIREEAPKXHL 627
Db	787	LQTLQKER--BDLLEETRSCKKIDEL-----EDALSE-----LK--KETSOXHHI 830
Qy	628	KQLVASALSHIAFNALDNDGDQFQDNKV--VVKTHNSYALADGEHIFIVDPQKLSSTI 685
Db	831	KQLEE-----DNNSNIEWYQNKIEALKKDYESVITSVDKQ--TDIEKLQYKV 876
Qy	686	VDLEK-----ILKLEG-----LDYSTLRQTT--- 707
Db	877	KSLEKEIEBKIRLHYNVMDETINDDSLRKELEKSKINLTDAYSQIKKEYKDLVETTSQS 936
Qy	708	--QT-----QFRDMTTAGKALL--SKSNLRQGEKQKELQEAQFFI--GRVLDKAIKAEK 757
Db	937	LQQTNSKLDESFKQFTNQIKNLTDERTSLE--DKISLLEKQMFNLNDELQKGMEXEK 994
Qy	758	ALVTTKATQKQLLERSINKAVLATYNNNAIK-----KANVKRLEKELDLTLGLVEGKGP 811
Db	995	ADFKKRI-----SILQNNKEVEAVKSEYESKLSKIQNDLQDT----- 1033
Qy	812	LAQATWQGVYLLKTPLEPYEYIGLVNYPFKSGKLIYALDMSDTIGE--GQKQAYGNPI 869
Db	1034	-IYANTAQNNY-----EQLQKHADVSKTITSELREQLHTYKQV 1071
Qy	870	--LNVDED-----NEGYHALAVATLADYEGLDIKTILNSKLSQLSIRQVPTAAVHR 919
Db		

Db	1072	KTMLSDRQLEVALKENEKSWSSQKESL--LEQDLDS---NSRIEDLSSQKVL---LYDQ 1123
Qy	920	AGIFQAIQNAAAEAEQLLPKPGTHS-----EKSSSSES 952
Db	1124	IQIVTA---ADKEVNNSTNGPGLNNILITLRRERDILDTKVTVAERDAKMLRQKISLMDV 1180
Qy	953	ANSKDRGLQSNPKTNRGHRSAIL 975
Db	1181	ELQDARTKLDNSRVEKENHSSII 1203
RESULT 62		
ADS43855		
ID	ADS43855	standard; protein; 1875 AA.
XX	AC	ADS43855;
XX	DT	02-DEC-2004 (first entry)
XX	DE	Bacterial polypeptide #22285.
XX	KW	Recombinant DNA construct; transformed plant; improved plant property;
XX	KW	cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;
XX	KW	pathogen tolerance; pest tolerance; plant disease resistance;
XX	KW	cell cycle pathway modification; plant growth regulator;
XX	KW	homologous recombination; seed oil yield; protein yield; carbohydrate;
XX	KW	nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX	KW	bacterial polypeptide.
OS		Bacteria.
FN	US2003233675-A1.	
XX	PD	18-DEC-2003.
XX	PF	20-FEB-2003; 2003US-00369493.
XX	PR	21-FEB-2002; 2002US-0360039P.
PA	(CAOY/) CAO Y.	
PA	(HINK/) HINKLE G J.	
PA	(SLAT/) SLATER S C.	
PA	(CHEN/) CHEN X.	
PA	(GOLD/) GOLDMAN B S.	
PI	Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS,	
XX	WPI; 2004-061375/06.	
XX	New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.	
PS	Claim 1; SEQ ID NO 22285; 122pp; English.	
CC	The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or	







compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1778 AA;

Query Match 3.8%; Score 195.5; DB 6; Length 1778;  
Best Local Similarity 20.2%; Pred. No. 0.01;  
Matches 209; Conservative 137; Mismatches 330; Indels 361; Gaps 50;

QY 44 GADYAESSGSKLKNINETSQVDDTVTLFSDKRTT--PEKIKD-NLAKGPREQELKAVT 100  
DB 27 GADETTVSEDTAVKTAASATGIESGTGSDDETAEPKAEKAEASKETTEKEKAKT 86  
QY 101 ENTESEKQI---TSGSQLESKESLSLNKTVPTSTSNWEICDFITKNTLV---GLSKSGV 154  
DB 87 EEPASNIKTEINTDKSQLKT-----SLKAAVPAGSTY-----NSLPPDNLAK--- 130  
QY 155 EKLQSDTHLVLPQAAAGTQIQVASFAPTPDKKTAIABYTRAGENGISQLDVGKEI 214  
DB 131 -KLAV---IITGNAATGNEVDASALL-----AISQLDL----- 161  
QY 215 INGEVFNNSYLLKKVTIPGYKHIGQDAFVNDKNIAEVLNPSLETISDYAFHLAKOI 274  
DB 162 --SGETGND-----PTDISNIEGLQYLE--NLTSNLSEN--NISDLA-----PLK-- 201  
QY 275 DLDPNLKAIKELGELAFFNQITKLSLPRQLMRLAERAFKSNHIKIE--FRGNSLKVIGA 332  
DB 202 ----DLVNLVSLNLSNRNLVNLGVEDLVNLQELNVSA--KALEDISQVASFVLKEI 255  
QY 333 SFQDNDLSQLMPDGLEKIESEAFNGPDGHHYNNRVVLWTKSGKNPSGLATENTYVNP 392  
DB 256 SAQGCNKTLEL-----KNPAGAVL----- 275  
QY 393 KSLWQSPEDITYKLEEDFTYKNSVTGFSNKGLOKVRKNKLEIP-----KQHG 444  
DB 276 -----PEL-----ETFYQENDLTNLTS--LAKLPKLNLYIKGNASLSKLETLNG 319  
QY 445 VT-----ITEIGD-----NAFRNV-DFQNKTLRKYDL 470  
DB 320 ATKLQILDASNCOTDLTGLDISGLEMLQISGCSKLEIYSLKLNPLNVLNITADSCAI 379  
QY 471 BEVKLPSTIRKIGAFQSN-NLKSFEASDDELEIK-----EGAFMNN-----RIETL 517  
DB 380 EDLGLTNLPLKQLTVLSDNENLTNTAITDLPQLKTLTLDGGGITSIGTLDNLPKLEKL 439  
QY 518 ELKDKLVITGDAAFHNHIIYAVLPSVQEIERSAFRQNGANNLIIPWG--SKVKTIGEMA 575  
DB 440 DLKE-----NQITSI---SEITDLPRLSYLDVSVNNLTITIGDKLPLLEWLN 484  
QY 576 FLNRLHLDLSBOKLTETIPVQAF--SDNALKEVLLPASLKTIREEAFKKNHLKQLBV 632  
DB 485 VSNR-----LSDVSTLTFFPSLNYINISNNVIRTVGKTELPSELKFEFAQNNSISDISM 539  
QY 633 ASALSHI-----AFNALDNDGDGFQFNKVVVTKHNSYALADGEHPIDVPDKLSSTIV-- 686  
DB 540 IHDPNLRKVDASNNLITNIG--TFDN-----LPKLQSLDVHNSRITSTSVIH 585  
QY 687 DLEKILKLEGLDYSTLRTOTQTFRDMTAGKALLSKSNLROGEKQKFLQEAQPFGLGRV 746  
DB 586 DLPSL-----ETFNAQTNLITNIG--TMDNLPD-----LTYVNLISFNRI 622

QY 747 -----DIDKAIKAEKALVTKKATKNGQLLERSINKAVLAYNNSAIKKNVYKRL 796  
DB 623 PSLAPIGDLPNLETLIVSDNNSYLSLGTMDG-----VFKLRILDIQ 664  
QY 797 KELDLLLTGLVEGKGLPAQATWQGVYLLKTLPLPEYVYGLNVPYDK-SG-----KLIYA 850  
DB 665 NNYLNTYCTEGNLSLSDLT-----NLTELNRNNVYIDDISGLSTLSRLIY- 711  
QY 851 LDMSDTIGEGQKDAYGNPLANVDEN-RGYHALA-----VATLADYEG 892  
DB 712 -----LNDSNKIEDISALSMLTNLQELTENNKIENISALSDLEN 752  
QY 893 LDIKTILSKLSQTSIRQVPTAAHYHRAGIFQA-----IQNAAAEQQLPKP 940  
DB 753 LNKLVVSKNIIDISPVANM---VNRGAIVTASQTVTLPTVLSVQSSTIDNPVWYD 808  
QY 941 GTHSEKSSSESANSKD 957  
DB 809 GTLLAPSSSIGNSGNYKD 825

RESULT 65  
ABP40235

ID ABP40235 standard; protein; 3696 AA.

XX AC ABP40235;

DT 24-JUL-2002 (first entry)

XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5080.

XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX KW antibacterial; gene therapy.

XX OS Staphylococcus epidermidis.

XX PN US6380370-B1.

XX PD 30-APR-2002.

XX PF 13-AUG-1998; 98US-00134001.

XX PR 14-AUG-1997; 97US-0055779P.

XX PR 08-NOV-1997; 97US-0064964P.

XX PA (GENO-) GENOME THERAPEUTICS CORP.

XX PI Doucette-Stamm LA, Bush D;

XX DR WPI; 2002-381255/41.

XX DR N-PSDB; ABN92780.

XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermis  
XX PS polypeptide, useful for diagnosing and treating bacterial infections.

XX PS Disclosure; SEQ ID NO 5080; 267pp; English.

XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
XX CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
XX CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have  
XX CC antibacterial activity and can be used in gene therapy. The sequences can  
XX CC also be used in the diagnosis and treatment of bacterial infections,  
XX CC particularly *S. epidermidis* infections. The sequences can be used to  
XX CC screen for compounds able to interfere with the *S. epidermidis* life cycle  
XX CC or inhibit *S. epidermidis* infection. N.B. The sequence data for this  
XX CC patent did not form part of the printed specification, but was obtained  
XX CC in electronic format directly from the USPTO web site

XX SQ Sequence 3696 AA;

Query Match 3.8%; Score 195; DB 5; Length 3696;  
Best Local Similarity 18.8%; Pred. No. 0.03;  
Matches 220; Conservative 164; Mismatches 478; Indels 310; Gaps 45;

QY 10 LTLTVSVVTHNOEVSLVKEPILKQTOASSISGADYAESGKSKLKINETSJGVPDVT 69  
Db 1082 ITHTVNVQKFOARQALIAKTNEKQSAINSNEGTT--IEEKQKAIQSLNDKXNLADQEI 1139  
QY 70 TDLFSDKR-----TPEKIKONLAK-----GPRQELKAVTENTSEKQ 108  
Db 1140 TQASNQNVNALNIGISNISKIOTFTKQOARDQVQKFOEKEAELNSTPHATQDEKQ 1199  
QY 109 ITSGSQSKESLSLNTKVPSTNWEICDFITKG-----143  
Db 1200 -DALTRLTQAKET-ALNDINOAQTNQNVDTALTSGIQNTQVNVVRKQBAKTTINDIV 1257  
QY 144 -----NTVLGLSKSGVEKLSQTDHLVLPQ-----AADGTQLIQV 178  
Db 1258 QOHKOSIONDDATTEEKAVANLVNASQOONY--ISKIDNATTNNQIDGIVSDGRQSN- 1314  
QY 179 ASFAFATP-----DKKTAIAEYTSRAGE-----NGEISOLDVDGKEII- 215  
Db 1315 ---ALTPDTSIKENAKNDIDIKAAOKKIKIQINDATDEEIQEANKKIEEAKIEAKNIQ 1371  
QY 216 -----NEGVFNSYLKKVTIPTGYKHIGQDAFVQNK-----NTAEVNLPSL 258  
Db 1372 RNRSDQVNEAKTNGINKNIENITPATTVKSEARQA-VQNKANEQINHIONTPDATNEEQ 1430  
QY 259 ETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSLPRQLM--RLAERAFKSHI 316  
Db 1431 EAINRVS-ABELARVOAQNAEHTQGVKTIKDDAITSLSRINAQVVEKESARNATEQKAT 1489  
QY 317 KTIETFRGSLKVGISAFODNLSQLMLPDGLEKIESEBAFTGNPDGDDHYNRNVLTQSG 376  
Db 1490 QQTQFINNDNATDEKEVANNLVATKQSLDNINSL-----SNDVENVAKVAG 1540  
QY 377 KN-----PSGLATENTYVNPDKSLMQSPESPEIDYTKWLEEDFTYQKNSVTFSGNKLQV 430  
Db 1541 INEIANVLPAVAKSKAKKIDIQKLAQIQNIQ-----THQ-TATTEKEAAIQLA 1590  
QY 431 KENKN--LEIPKQNGVTITIGDAPRNVDFQNKTLKYDLEBVKLPSTIRKIGA--- 484  
Db 1591 NOKSNEARFATIQNEHSNNGVAQKANGTHEI-----ELVMPDPAHKKSDAKQS 1637  
QY 485 ----FAFOSNNLKSPEASDLEIEIKEGAFMNNRIETLELKDKLVTIGDAAFHINHIAIV 540  
Db 1638 INDKNEQSNNTINT--TPDATDEEKQAL-----DKLKTAKAGY--NKVDAQ 1682  
QY 541 LPE\$VOEIGRSAPRONGANNLIFMGSKVKTIGEMAFPLNRLHLDLSEKQQLTEIPVOAF 600  
Db 1683 TNQQVSDAKTEAI--DTITNI--QANVAKKPSARVELDSKFEDL-----KRQINATP--- 1730  
QY 601 SDNALKEVLLPASLKTIREEAPKKNHLKQLEVASALSHIAPNALDDNDGDEQFNKVVK 660  
Db 1731 --NATEEE-----KQDAIQRLNGRDEKVNILN-----QQRDNVEV--- 1764  
QY 661 THNNGVALADGHEFVDPDKLSSTIVDL-----EKILKIRGL 698  
Db 1765 EQKNIGIGLEETIHANPRKSDALQELQTKFISOTELINNNKDATNEEKDEAKLLIS 1824  
QY 699 DYSTLRQTTQTO-----FRDMTITAGKALLSKNSURQGEKQKFLQBAQFPLGR- 745  
Db 1825 KXKTTINQAQTNQVONAKONGMNEIATITPATITKDTAKTADKKAQOQVTTINGNN 1884  
QY 746 --VLDLKAIAK--AEKALVTKATKNGQLLESINKA-----VLAYNN---SAIKKANVR 794  
Db 1885 DATDEKAEARKLVEKAKIRKASNTINSDEEVNGAKTNGLEKINNIQPSQTQTKNAQ 1944  
QY 795 -----LEKELDILT-----GLVEGKGPLAQATMVQGVLLKTLPL 830  
Db 1945 EINDKAQEQLIQINNTPDATDEEKQEAETNRVNAQLAQAIQNNASHSTQEVNESKTNIA 2004  
QY 831 PYYITGLNVYFKSKLIALVDMSTIGEGQKDAVGNPILNVDBNEGVHALAVATLAD- 889  
Db 2005 TTKSVQPNVI-----KKPTAINLQTEANNQKTLIGNDGNATDDEKAAQLVTKLNEQ 2059

QY 890 ----YEGLDIKTILNSKLSQLTSTIROVPTAYHAGIFQAIONAA--AAEQLLPKPCTH 943  
Db 2060 IQKHESQDNQVNVKQAQAITAIKLINAHKQDQDAINILNLAEKSKSDIRANQDATT 2119  
QY 944 SEKSSSESANSKDRGLQSNPKTKNRGRHSAIL 975  
Db 2120 EEKNTAQSID--DTLAQARNNINGANTNALV 2149  
RESULT 66  
ADS06647  
ID ADS06647 standard; protein; 3696 AA.  
XX ADS06647;  
AC ADS06647;  
XX 04-NOV-2004 (first entry)  
XX Staphylococcus epidermis polypeptide seqid 5942.  
DE antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;  
KW recombinant expression vector; infection; computer readable medium;  
KW computer based system.  
XX Staphylococcus epidermidis.  
OS US2004147734-A1.  
FN 29-JUL-2004.  
XX 01-DEC-2003; 2003US-00724972.  
XX 08-NOV-1997; 97US-0064964P.  
PR 13-AUG-1998; 98US-00134001.  
PR 29-NOV-1999; 99US-00450969.  
XX (DOUC/) DOUCETTE-STAMM L.  
PA (BUSH/) BUSH D.  
XX Doucette-Stamm L, Bush D;  
PI WPI; 2004-580138/56.  
XX N-PSDB; ADS02875.  
DR New isolated polypeptide and encoding nucleic acid derived from  
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or  
PT treating an S. epidermidis bacterial infection.  
XX Claim 17; SEQ ID NO 5942; 741pp; English.  
XX The invention describes an isolated nucleic acid comprising a nucleotide  
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:  
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any  
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as  
CC given in the specification. Also described are: a recombinant expression  
CC vector; a cell comprising a recombinant expression vector of (1);  
CC producing an S. epidermidis polypeptide; an isolated nucleic acid  
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a  
CC vaccine composition for prevention or treatment of an S. epidermidis  
CC infection, comprising a nucleic acid cited above and a carrier; treating  
CC a subject for S. epidermidis infection; a recombinant or substantially  
CC pure preparation of an S. epidermidis polypeptide or its fragment; a  
CC vaccine composition for prevention or treatment of an S. epidermidis  
CC infection; detecting the presence of a Staphylococcus nucleic acid in a  
CC sample; a computer readable medium having recorded in it the nucleotide  
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based  
CC system for identifying fragments of the Staphylococcus genome of  
CC commercial importance; a computer based system for identifying fragments  
CC of the Staphylococcus plasmids of commercial importance; identifying  
CC commercially important nucleic acid fragments of the Staphylococcus  
CC genome and/or plasmids; and identifying an expression modulating fragment  
CC of the Staphylococcus genome and/or plasmids. The methods and  
CC compositions of the present invention are useful for the diagnosis,  
CC prevention and/or treatment of an Staphylococcus epidermidis bacterial

CC infection. This is the amino acid sequence of a S. epidermidis protein of  
CC the invention.

Sequence 3696 AA:

Query Match 3.8%; Score 195; DB 8; Length 3696;  
Best Local Similarity 18.8%; pred. No. 0.03;  
Matches 220; Conservative 164; Mismatches 478; Indels 31

Qy	10	LTLTTSVTVHTNQVFLSVKEPIILUKQTAQSSISIGADYAESSGSKSLKUKINETSGPVDVT	69
Db	1082	ITHTNVNVQKQPQARQALIAKTNEKSGAINSDNEGTT--IBEKQKAIQSLDANKLADAEQI	1139
Qy	70	TDLPSDKR-----TTPEKIKDNLAK-----GPREQELKAVTENTSEKQ	108
Db	1140	TQAMSNQNVNDNALMIGISNISIKIQTNFTTKQQAQDQVNVQKFPQKEAELNSTPHATODEKQ	1199
Qy	109	ITSQSQLEQSEKISLKNKTPSTSNWEICDPIKTG-----	143
Db	1200	-DALTRLTQAKET--ALANDINQAQTNQNVDPALTSGIQNIQNTQVNVTKQOEAKTTINDIV	1257
Qy	144	-----NTVLUGSKSGVEKLSQTDHLVLPQSQ-----AAGTQLIQV	178
Db	1258	QOHKQSIQNNDATTTEEKEVANLNVASQQNV--ISKIDNATTNNQIDGIVSDGRQSIN-	1314
Qy	179	ASFAFTP-----DKKTAAEYTSRAGE-----NGEISQLDVGKEI-	215
Db	1315	---AITPDTSIKRNAKNNDIDIKAADKKIKIQRINDATDEIEQANRKIEBAKTEAKDNQI	1371
Qy	216	-----NEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNK-----NIAEVNLPESL	258
Db	1372	RNSTRDQVNEAKTINGINKIENITPATTVKSEARQA--VONKANEQINHIQNTPDATNEEQQ	1430
Qy	259	ETTSIDYAPAHALAKQIDLPNLKAIGELAFPDNQITGKLSLPQLM--RAEERAFKSNHI	316
Db	1431	EAINRVS--AELARVQAQINAEHTTQGVKTIKDADAITSLSRINAQVVEKESARNAIEQAT	1489
Qy	317	KTIEFRGNSLKVIGEASFQDNDSLQMLPDGLEKIESEAFPTGNPGDDHYNNRVVLMTKSG	376
Db	1490	QOTQFINNNDNATDEEKEVANLNVIAATKQSKLWNISLS-----SNNDVENAKVAG	1540
Qy	377	KN-----PSGLAPENTYVNPDKSLWQSEPIDYTKWLEEDFTYQKNSVTGFSNKGLOKV	430
Db	1541	INBIANVLPATAVKSKAKKDIDOKLAQIINQIQ-----THQ--TATTEEKAAIQLA	1590
Qy	431	KRNKN-----LEIPKQHNQVTTTEIGDNAPRVNDPONTLYRKVDUEVKLPSTIRKIGA---	484
Db	1591	NOKSNEARTAIQNEHSNNGVAQAQKSGNIGHEI-----ELVMPDAHKKSDAKQS	1637
Qy	485	----FAFQSNLKLKSFASDDLLEETKEGAFWNRRITLLELKDGLVTIGDAAPHNHIYAIV	540
Db	1638	IDNKYNEQSNFTINT--TPDATDEEKQAL-----DKUKIAKDAGY--NKVDQQAQ	1682
Qy	541	LPHSVQBQIGRSAPRQNGANNLIFMGSKVKTLIGEMAFLSNRLEHLIDSEQQLTEIPVQAF	600
Db	1683	TNQOVSDAKTEAI--DIITWI--QANVAKKPSARVELDSKFEEL-----KQINATP----	1730
Qy	601	SDNALKEVLLPASLUKTTREERAFKKNHUKQLEVASALSASHIAPNALDDNDGDEQFNDKVVVK	660
Db	1731	--NATEBE-----KQDAIORLNGKRDEVRKNLIN-----QDRRDNEY----	1764
Qy	661	THNSYALADGEHFI VDPDKLSSTVIDL-----EKILKLEGL	698
Db	1765	EQHKNI GLQLETHIHPNRKSDALQELQTKPTISQTELINNNKDATNEEKDEAKRLLLEIS	1824
Qy	699	DYSTLRQTTQIQ-----FRDWTTAGKALLSKSNLRQGEKQKFLQEAQFFLGR--	745
Db	1825	KNKTIITINQAQTNQVNDNAKDNQWNEIATIIIPATTIKTDAKTADKKAQEQVITIINGNN	1884
Qy	746	--VOLDKAIK--AEKALVTYKATYNGQLLERSINKA-----VLAYNN-----SAIKKANVXR	794
Db	1885	DATDEEKAERKLVKEAKIEAKSNITNSDTEREVNGAKTNGLEKINNIQSTOTKTNAKQ	1944

RESULT 67  
AAU35908

AAU35908	
ID	AAU35908 standard; protein; 949 AA.
XX	
XX	
AC	AAU35908;
XX	
XX	
DT	14-FEB-2002 (first entry)
XX	
XX	
DE	Helicobacter pylori cellular proliferation protein #221.
XX	
XX	
KW	Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW	antibacterial; drug design.
XX	
XX	
OS	Helicobacter pylori.
XX	
PN	WO200170955-A2.

XX  
PD 27-SEP-2001.

XX  
PF 21-MAR-2001:

XX  
PR 21-MAR-2000: 2000US-0191078P:  
ET 12TH 2001; FZCING 0500051000

PR 26-MAY-2000; 2000US-0207727P  
PR 23-MAY-2000; 2000US-0206848P.  
PR 21-MAY-2000; 2000US-0151070E

FR 26-MAY-2000; 2000US-0207127E  
PR 23-OCT-2000; 2000US-0242578P  
PR 27-NOV-2000 2000US-0253625P

PR 27-NOV-2000; 2000US-0253625P.  
PR 22-DEC-2000; 2000US-0257931P.  
DE 15-FEB-2001; 2001US-0269308P.

PR 16-FEB-2001; 2001US-0269308P.  
XX  
D3 (P17E) P17E23 DUADRM INC

PA (ELIT-) ELITRA PHARM INC.  
XX  
BT

PI Haselbeck R, Ohlisen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;

XX  
DR WPI; 2001-611495/70.

DR N-PSDB; AAS53767.  
XX

PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids.

PS Example 3: SEQ ID NO 11501; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the genes,  
CC their use in the discovery of novel antibiotics, the essential genes  
CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*  
CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,  
CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also  
CC useful for the identification of potential new targets for antibiotic  
CC development. The antisense nucleic acids can also be used to identify  
CC proteins used in proliferation, to express these proteins, and to obtain  
CC antibodies capable of binding to the expressed proteins. The proteins can  
CC be used to screen compounds in rational drug discovery programmes. The  
CC antisense nucleic acid sequence is also useful to screen for homologous  
CC nucleic acids which are required for cell proliferation in a wide variety  
CC of organisms. The present sequence represents an essential prokaryotic

CC cellular proliferation protein. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 949 AA;

Query Match 3.8%; Score 194; DB 4; Length 949;

Best Local Similarity 19.6%; Pred. No. 0.0052;

Matches 202; Conservative 150; Mismatches 343; Indels 338; Gaps 48;

QY 6 KTVALTLLTVSVVTHNQ--EVSFLVKEPILKCTQASSISGADYAESGSKLKINETSG 63  
DB 29 KDIGLELKNFMWTEQAGLKYIVDGIKEQIQAN-----OPTKN 69  
QY 64 PVDDTVTDLFSDKRTTPEKIDKNLAKPREQELKAVTENTES-EKQITSGSQLEQSKESL 122  
DB 70 PQDNKDDL--NTAATPKPLAKKASKTPKKEETKAQPKPKTKKKEAPAPIIKKEIE 127  
QY 123 SLANKTVPTSNWEICDFITKGNLTVGLSKSGVEKLSQTDHLVLPQADGTQLIQVASFA 182  
DB 128 IVNTFENQPLVE-----NTPKAVSHSQIEKAKQLQEIQKSGREA-----LNKLQTSN 175  
QY 183 FPDCKTAIAEYTSRAGENGEISQLDVDGKEIINEGEVFNYSLLKKVVTIPTGYKHIGODA 242  
DB 176 TWTNNANSASNVNAKK--EISEVKKQEQEIKRH-----ENIKRTGFRVIRKND 224  
QY 243 FVDNKNIAFVNLPESLETISDYAFALHALKQIDLPNLKAIGELAFQDNQITGKLSLPRQ 302  
DB 225 --ETENETENSVTESKPKTQSAAIPEDIKK-----EWQEKDKQETKTKPKSK 271  
QY 303 LMLAERAFKSHIYTI-----EPRGNSLVKIGEASQDNLDLSQML-----PGLSKI 351  
DB 272 --PKATPTAKNNKSHKIDFSDVRDFKNDI-----YDDTDEILLFDLHEQDNLANKE 321  
QY 352 ESEAFTPGNPDHYNRVVLTMTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEE- 410  
DB 322 EEE-----KEARQINDRVV---QRKNP-----WMNEA 347  
QY 411 DFTYQKNSVTGSNKGLOKVRKNKLEIPKQHNGVTITEIGDNAFRNV----- 458  
DB 348 GIKRQSKKRVPRNDSQKVIQSA-IAIPEE--VRVVEFAQKANLADVIKTLFNLGL 403  
QY 459 -----DFQNKTLRKVDLSEVKLPSTIRKIGAFQSNLKSFEASDDLEBEKE----- 506  
DB 404 MYTKNDFLDKDSIBTIAEFHLEISVQ-----NTLEEFVEVEVLEGVKKERPPVVT 454  
QY 507 --GAFMNNRIETLE-LKDKLVTIGDAAFHINHIYAIVLPE-----VQETGRSAF--R 554  
DB 455 INGHVDHGKTSLLDKIRKRVVAHTAGGITQHIGAYVKEKNKNWSPIDTPGHEAFSOMR 514  
QY 555 QNGAN-----NLIFM-----GSKVKTILGEM-----AFLSNRLEHLDLSEQKOLTE-- 594  
DB 515 NRGAVTDAIVIAAADGVKQOTIEALEHAKAANVPVIFAMKNKMDKENVNPKLKAESA 574  
QY 595 -----IPVQAFS-----DNALKEVLIPASLKTIRERAPKKNHLKQLEVASA 635  
DB 575 ELGYNPVDWGGHEFIPVSATKGTGDIDNLLLETILIQADIM-----ELKATIEGSA 624  
QY 636 LSHAFNALDDNDGDEQFNDKNVVKTHNSYALADGHEFIVD-----PDKLSSTIV 686  
DB 625 RAVVLEGSVEKRG-----AVATVIVQSGLTSLVGDSPFAETARGKVRMTDDGKSIQ 677  
QY 687 DLE-KILKLIBGLDYSLTRQTQTOTFRDMTTAGKALLSKSN-----LRQGEKQKFLQBAQ 740  
DB 678 NLKPSMVALITGLS-----EVPVAGSVLIGVENDSIARLQAKREATYLRQ-- 722  
QY 741 PFLGRVDLDKATAKAKEA-----LVTKKATKN-----QOLLERSINKAVLAYN 783  
DB 723 -----KALSKSTKTVSFDELSEMVANKELKNIPVIKADTQGLS--BAIKNSLLEIN 771  
QY 784 NSAIKKANVGR-----LEKELDLL-----TGLVEGKGPLAQATMVQGVYL 823

DB 772 NEEVAIQVHSGVGGITENDLSLVSSSEHAVILGFNIRPTGNVKNKA----- 818  
QY 824 LKTPLPPEYVIGLVNVPYFDKSGKLIYALDMSDTTIGEGQKD---AYGNPILNVD-----ED 875  
DB 819 -----KEYNVSIKTY-----TWIYALI-----EGMRSLLGLMSPIEEEHTGQAEV 860  
QY 876 NEGYHALAVATLA 888  
DB 861 RETFNIPKVGTTIA 873

RESULT 68

ADH62806  
ID ADH62806 standard; protein; 1182 AA.

AC ADH62806;

XX 15-APR-2004 (first entry)

XX Lactobacillus johnsonii mucin binding protein sequence SeqID24.  
DE mucin binding; probiotic; glycoprotein; epithelial cell;  
KW gastrointestinal tract; lung; uterine cervix.

XX Lactobacillus johnsonii.

XX EPI382970-A1.

XX 21-JAN-2004.

XX 15-JUL-2002; 2002EP-00015609.

XX 15-JUL-2002; 2002EP-00015609.

XX (NEST ) SOC PROD NESTLE SA.

XX Pridmore RD;

XX WPI; 2004-111519/12.

XX N-PSDB; ADH62786.

XX Novel mucin binding polypeptide encoded by novel gene of Lactobacillus  
johnsonii, useful for binding bacteria to mucin.

XX Claim 8; SEQ ID NO 24; 225pp; English.

XX This invention relates to novel mucin binding polypeptides and the DNA  
sequences which encode them. In particular, the genes are derived from  
Lactobacillus johnsonii. The invention may allow a better understanding  
of the mechanism by which probiotics (such as L johnsonii) can bind to  
mucins, large glycoproteins present on the surface of a large number of  
epithelial cells including those in the gastrointestinal tract, the lung  
or the uterine cervix. The present sequence is that of an L johnsonii  
mucin binding protein of the invention.

XX Sequence 1182 AA;

Query Match 3.8%; Score 194; DB 8; Length 1182;

Best Local Similarity 19.7%; Pred. No. 0.0071;

Matches 222; Conservative 174; Mismatches 421; Indels 308; Gaps 53;

QY 48 AESSGSKLKINETSGPV-DDTVTDLFSDKRTTPEKIDKNLAKG-----PRQELK 97  
DB 84 SSSDGYQKIEV-----FLSDQVKTLRAGKNDLAIADFNDATNAGTSGTSNKPEINFG 138  
QY 98 AVTENTESEKQITSGSQLEQSKESLSLKNKTVPSNSNEICD-----F 139  
DB 139 LIIDNNLPQKITVSDGDYDMTDDSYTISGTYPKVTYTDKQKHEHDLNISYDEASERF 198  
QY 140 ITKGNLTVGLSKSGVEKLSQTDHLVLPQADGTQLIQVASFAFTPKKTAIA---EYTS 196  
DB 199 VTKLPVLSVDYDINVKVFADBEHETLITQKRNIVSLV-----PPKLESLSKVDQDQTY 250





```
Db 1381 -EVAQALAKEKEKALAA-IDQAQTSQVNOQAATNGVSAIK-----LIQETKVKPA 1430
Qy 234 GYKHIGQDAFVNDKNAIAEVLNPE-----SLETISDYAPAHAL-----KOID 275
Db 1431 AREKINOKA---NELRAKINQDKEATABEROVALDKINE--FVNOAMTDITNRTNQOVD 1485
Qy 276 LPDNLKATIGELAFDQNTGKLSLPRQLMLAER-----AFKSNHIKTIEFRGNSLKV 328
Db 1486 -DTSQALDSIA-----LVTPDHIVRAAADAVKQVEAKKREIEQAHEHATDEBKQ 1535
Qy 329 IGEASFQDND-----LSQLMLPDGLEKIESAFTGNPGDDHYNRVVLTWTKSGKNPSGL 382
Db 1536 VALNQLANNEKRALQNDQAIANNQVKTETNGIATLKG---VQPHIVI-----KPEAQQ 1587
Qy 383 ATENTYVNPDKSLWQESP-----EIDYTKWL-----EEDFTYQKNSVTGFSN--- 424
Db 1588 AIKASAENQVESI-KDTPHATVDELDEANQLISDTLKOQQOEIENTNQDAAVTVRNTQTI 1646
Qy 425 KGLQKVK-----RNKN-----LEIPKQHNGVTITEIGD--NAFRNVD 459
Db 1647 KAIEQIKPKVRKRAALDSIENNNKQLDAIRNTLDTQDERDVAIDTLNKIVNTIRKDI 1706
Qy 460 FQNKTLRYDLBE-----VKL-----PSTIRKIGAFAPQSNLKSFEASD-DLBEI 504
Db 1707 AQNKINAEDVRETETDGNNDNIKVILPKVQVKPAARQSVGVKABQAQAL--IDQSDLSITEE 1764
Qy 505 KEGA-----FNNRIETLELKDVLITGDAAAPHINHIYVLP-----ESVQBIG 549
Db 1765 RLAAKHLVEQALNQAIQDINHADKTAQVNOQSINAQNIISKIPATTVKATALQIQNIA 1824
Qy 550 RAFPQNGANN-----LI FMGSKVK---TLGMAFL-----SNRLEHLD 585
Db 1825 TNKINLIKANNEATBEONIAIAQVEKELIKAKQOIASAVTNADVAYLLHDEKNEIRIE 1884
Qy 586 -----LSEQKQLTIPVQAFSDNALKEVLLPASLKTIREAFKKNHLKQLEVASALSHI 639
Db 1885 PVINRKASAREQLTWL-----FND---KKQAIENAIQATVEE--RNSILAQLQ--NIYDT 1932
Qy 640 APNALDDNDGDBQFON--KVVYKTHNSYALADGSHFIVDDPKLSTTIVDLEKILKLIEG 697
Db 1933 AIGQIDQDRSNAQVDKTASLNLQTIHDL-----DVHPICKPDPAEKTINDDLARVTALVQ- 1986
Qy 698 LDYSTLRQTQTPQFDMITAGKALLSKSNLRQGEKQFLOEAQFPLGRVLDLKATAKAEK 757
Db 1987 -----NYRKVSNRNK-----ADALK 2001
Qy 758 ALVAKKATNGQLLERSIN--KAVLAYNNSAIKKANVKRLEKELDLTLGLVEGKGLAQ 814
Db 2002 AITALKQWDEELKTARTNADVDVILKRFNVALSDIEAVITEKNSLLR--IDN---IAQ 2056
Qy 815 ATMQGVYLLKTPPLPEYIGLVNYVFKSGKLIYALDMSDTIGBGQKDAYGNPILNVDE 874
Db 2057 QT-A---YAKFRAIATPEQLAKVKVLDIQ-----YVAD-----GNRW--IDE 2091
Qy 875 DNEGYHALAVATLADYEGLDIKTLNSKLSQTSIROVPTAAYHRAGIFQALQNAABAE 934
Db 2092 D-----ATLN-----DIKQHTQFIVDEILAIK-LPAEA-----TKVSPKEI 2126
Qy 935 QLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG 982
Db 2127 QPAPKVCPTPIKKEETHSEKVEKE-----LPNTGSEG 2158

RESULT 70
ID ABM71190
XX ABM71190 standard; protein; 2186 AA.
AC ABM71190;
XX
XX
DT 20-NOV-2003 (first entry)
XX
DE Staphylococcus aureus protein #430.
XX
```

```
KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;
KW enzymatic assay; antibiotic target.
OS Staphylococcus aureus.
XX WO200294868-A2.
PN 28-NOV-2002.
XX 27-MAR-2002; 2002WO-IB002637.
XX 27-MAR-2001; 2001GB-00007661.
PA (CHIR-) CHIRON SPA.
XX Masignani V, Mora M, Scarselli M;
XX MPI, 2003-120786/11.
XX N-PSDB; ACF72750.
DR New Staphylococcus aureus protein, useful as a vaccine for treating or
PT preventing Staphylococcal infection, specifically an infection caused by
PT S. aureus, e.g. sepsis.
XX Claim 1; SEQ ID NO 860; 49pp; English.
XX
CC The invention relates to novel genes and encoded proteins from
CC Staphylococcus aureus. A composition comprising the S. aureus protein, a
CC nucleic acid encoding the protein, or an antibody to the protein, is
CC useful as a pharmaceutical, particularly as a vaccine for treating or
CC preventing infection due to Staphylococcus bacteria, specifically an
CC infection caused by S. aureus. The composition is particularly useful for
CC treating or preventing sepsis in a patient. The composition can also be
CC used for diagnostics. The protein is also used in an assay for enzymatic
CC studies and as a target for antibiotics. This sequence represents one of
CC the novel S. aureus proteins of the invention
XX
SQ Sequence 2186 AA;
Query Match 3.8%; Score 194; DB 6; Length 2186;
Best Local Similarity 19.1%; Pred. No. 0.017;
Matches 216; Conservative 183; Mismatches 387; Indels 342; Gaps 58;
Qy 20 HNOEVSLVKEPIIKQTQASSISGADYAESGSGSKLKNETSPGVDVTVDLFSDFD----- 75
Db 1208 YNAKLAEINATPDATNDEKNAINTLNQDROQAATESIKQANTNAEVDQAATVAENNIDAV 1267
Qy 76 -----KRTTEKIDKNLAKGPREQLKAVTENTESEKQITSGSOLESK-----SLSLN 125
Db 1268 QDVTVVKQAARDKITAIEVAK--RIEAVKQTPNATDEBKQ-AAVNIQNLKQQAQINQINQN 1324
Qy 126 KT---VPSTSNWEI-----CDPITKGNTLVGLSKSGVEKLSQTDHLVLPQAADGTQL 175
Db 1325 QTNQDQVTTTNAQVNAINDNVEAEVVIKTAADIEKAVEKQKQIDNSL-----DSTDNEK- 1380
Qy 176 IQVASFAPTPDKTAAIABYTSRAGENGEISOLDVDGKEIINEGEVFNYSLLKKVT--IPT 233
Db 1381 -EVAQALAKEKEKALAA-IDQAQTSQVNOQAATNGVSAIK-----LIQETKVKPA 1430
Qy 234 GYKHIGQDAFVNDKNAIAEVLNPE-----SLETISDYAPAHAL-----KOID 275
Db 1431 AREKINOKA---NELRAKINQDKEATABEROVALDKINE--FVNOAMTDITNRTNQOVD 1485
Qy 276 LPDNLKATIGELAFDQNTGKLSLPRQLMLAER-----AFKSNHIKTIEFRGNSLKV 328
Db 1486 -DTSQALDSIA-----LVTPDHIVRAAADAVKQVEAKKREIEQAHEHATDEBKQ 1535
Qy 329 IGEASFQDND-----LSQLMLPDGLEKIESAFTGNPGDDHYNRVVLTWTKSGKNPSGL 382
Db 1536 VALNQLANNEKRALQNDQAIANNQVKTETNGIATLKG---VQPHIVI-----KPEAQQ 1587
Qy 383 ATENTYVNPDKSLWQESP-----EIDYTKWL-----EEDFTYQKNSVTGFSN--- 424
```

Db 1588 AAKASAEQVESI-KDTPHATVDELDEANQLISDTLKAQAEIENQNDAAVTDVRNQT 1646  
 QY 425 KGLQKVK-----RNKN-----LEIPKOHNGVITTEIGD--NAPRND 459  
 Db 1647 KALEQIKPVRRKBAALDSIENKNQOLDAIRNTLDTQDERDVALDTINKVINTKNDI 1706  
 QY 460 FQNTURKYDLEB-----VKL-----PSTIRKIGAPAFQSNLKSFEASD-DLEBI 504  
 Db 1707 AQNKTAENVDRTEGTGNDINIKVILPKVQVQKPAARQSVGKABEAQNAL--IDQSDLSTEBE 1764  
 QY 505 KEGA-----FWMNRLETLKDLKLVIGDAAPHINHIYALVP-----ESVQETG 549  
 Db 1765 RLAAKHLVQALNQAIQDINHAKDTAQNQDSINAQNIISKIKPATVTKATALQIQNTA 1824  
 QY 550 RSAFRONGANN-----LIFMGSVKV-----TLCEMAFL-----SNRLEHLD 585  
 Db 1825 TNKINLIKANNEATDEQNIATAQVEKELIKAKQOIASAVTNADVAYLHDEKNEIREIE 1884  
 QY 586 -----LSQOKQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKQVHLKQLEVASALSHI 639  
 Db 1885 PVINRKASAREQLTTL-----FND-----KQQAIEANIQATVEE--RNSILAQLQ---NIYDT 1932  
 QY 640 AFNALDNDGDQFON--KVVVKTTHNSYALADGEHFIVDPDKLSSTIVDLEKILKLIBG 697  
 Db 1933 AIGQIDQDRSNAQVDKTASINLQTIHDL-----DVHPKPKPAEKTINDDLARVTALVQ- 1986  
 QY 698 LQYSTLRQTQTOFRDMTGTAGKALLSKNSLROGEKQKFLQEAQFFLGRVDLDKAIKAEK 757  
 Db 1987 -----NYRKVSNRK-----ADALK 2001  
 QY 758 ALVTKATKNGOLLERSIN---KAVLAYNSAIAKKANVRLEKELDLLTGLVEGKPLAQ 814  
 Db 2002 AITALKQMDDEELKTARTNADVAULKRFNVALSDIEAVITEKENSLLR--IDN---IAQ 2056  
 QY 815 ATWQGVYLLKPLPEYVYIGLVYFDKSGKLIYALDMSDITIGEQKDAYGNPILNVD 874  
 Db 2057 QT-----YAKFKAIAATPEQLAKYKVLIDQ-----YVAD-----GNRM--IDE 2091  
 QY 875 DNEGVHALAVATLADYEGDLDTILNSKLSQTSIROVPTAAVHRAGIFQAIONAAEAE 934  
 Db 2092 D-----ATLN-----DIKQHTQFIVDEILAIAK-LPAEA-----TKVSPKEI 2126  
 QY 935 QLLPKPGTHSEKSSSEANSKDRGLQSLQPNKPTRGRHSAILPRTGSKG 982  
 Db 2127 QPAPKVCTPIKKEETHESRKVEK-----LPTGSEG 2158

RESULT 71  
 ADH62805  
 ID ADH62805 standard; protein; 2209 AA.  
 XX AC ADH62805;  
 XX DT 15-APR-2004 (first entry)  
 XX DE Lactobacillus johnsonii mucin binding protein sequence SeqID23.  
 XX KW mucin binding; probiotic; glycoprotein; epithelial cell;  
 XX KW gastrointestinal tract; lung; uterine cervix.  
 XX OS Lactobacillus johnsonii.  
 XX PN EP1382970-A1.  
 XX PD 21-JAN-2004.  
 XX PF 15-JUL-2002; 2002EP-00015609.  
 XX PR 15-JUL-2002; 2002EP-00015609.  
 XX PA (NEST ) SOC PROD NESTLE SA.  
 XX PI Pridmore RD;

XX WPI; 2004-1111519/12.  
 DR N-PSDB; ADH62785.  
 XX Novel mucin binding polypeptide encoded by novel gene of Lactobacillus  
 PT johnsonii, useful for binding bacteria to mucin.  
 XX Claim 8; SEQ ID NO 23; 225pp; English.  
 XX This invention relates to novel mucin binding polypeptides and the DNA  
 CC sequences which encode them. In particular, the genes are derived from  
 CC Lactobacillus johnsonii. The invention may allow a better understanding  
 CC of the mechanism by which probiotics (such as L johnsonii) can bind to  
 CC mucins, large glycoproteins present on the surface of a large number of  
 CC epithelial cells including those in the gastrointestinal tract, the lung  
 CC or the uterine cervix. The present sequence is that of an L johnsonii  
 CC mucin binding protein of the invention.  
 XX SQ Sequence 2209 AA;  
 Query Match 3.8%; Score 194; DB 8; Length 2209;  
 Best Local Similarity 19.7%; Pred No. 0.017;  
 Matches 222; Conservative 174; Mismatches 421; Indels 308; Gaps 53;  
 QY 48 AESSGSKLKNETSQPV-DTPTDLPSDKRTTPEKIKDNLAGK-----PREQELK 97  
 Db 1050 SSSDGVQKIEV-----PLSDEQVKTLAGKNDLATAVFDNATNAGTSGTSNKPGEIFG 1104  
 QY 98 AVTENTESEKQITSGSQLEQSKESLSLNKTVPTSTNWEICD-----F 139  
 Db 1105 LIIDNNLPQKITTVSDGYDMDTDSYITSGTYPEKYGYTDRKQKEHDLNLSYDEASERF 1164  
 QY 140 ITKGNLTGLSKSGVKLSQTDHLVLPQQAADGTOLIQVAFAPDPKKTAIA---EYTS 196  
 Db 1165 VTKPLPSVSDVDNTVKFYADEHETLITQKRINVSLV-----PPKESLKVDDQET 1216  
 QY 197 RAGENGESQLDVGDKKEIINE-----GEVFNYSLLK---KVTIPTGYKHIG 239  
 Db 1217 TGNEEAKLSQTSSEDTVEVSGKVSDDTDKVAVKVAGTKYSKPTKEHTFKVKVPSVGENT 1276  
 QY 240 QDAFVDNKNIARVNPESLETISDYAFALHAKQIDLPDLNK-----AIGELAFDQNT 294  
 Db 1277 MNVILTDKDGSSSVKQIVKS--SDRGKTVWSAKDVTDFNGIKFGTTSVNTETENYDPK-T 1334  
 QY 295 GKLSIPROLMLAERAFKSNHIKTIEFRGNSLKVIGEASFQNDLSQMLPDLGKEISE 354  
 Db 1335 KGLTLTGKVR-----PTTVIRIGDHTVKKVADGTF-----KLVL--DLGKHGAK 1377  
 QY 355 APTGNPGD---DHYNRRVLMTKSGKNPSGLATENTYVNPDKSLWQSPESPIDYTKWLEE 410  
 Db 1378 VPPVLIGDTTVNDTVQERLTFYVDSN-----NPELTNLQEKQDSGY----- 1418  
 QY 411 DFTYQKNSVTGFSNKGLOKVRKNKMLEIPKQINGVTIITEIGDNAPRNVDFO--NKLTKYD 469  
 Db 1419 -----VPVYTNKEEFKLQGTISDDYPYYSLLINDNNVDAN--WDDVDYNGNKNLKKSF 1469  
 QY 470 LEEVKLPSTIRKIGAFQSNLKSFEASDLEEI-----KEGAFNNRIETLELKD 521  
 Db 1470 SHSVKL-----KEGKNTF---NVVVVDNNDNRSEVQTLVYVYKKAQKIASPOLITATTASD 1521  
 QY 522 -KLVTIGDAAPHINHIYAI-----VLPE---SVQEIGRSAFR--QNGANNLIWGSKV 568  
 Db 1522 KKSVTVTGKADGNVLYSTDNGNKNVLPEDGVTYKNGKLLFKTVTKVGNSESEVVEYDV 1581  
 QY 569 KTLGEMAFLSNRLEHLDSEKQLTEIPVQAFSDNALKEVLLPA-----S 613  
 Db 1582 KTIGK-----EESTVDKSVQAQKQD--LRKLDQARALGNTGYTHESAKK 1625  
 QY 614 LKTIREEAFK---KN-HLKQLEVAS-----ALSHIAFNALD---DNDGEQDFDNKVVVK 660  
 Db 1626 LAQAQFEASKALKDKKNATLQELKQASEQLEEAIKNLVEKFPVDQNKDKVEDKOSQNAL 1685  
 QY 661 THNSYALADGHEFIVDPDKLSSTIVDLEKILKIEGLDYSTL-----RQTTQTQPRD--- 713

Db 1686 KEKLEETVKAGKGF--DKDKYTDDSV-----KVTKALDEAKVVLANKDANSTVDQDAID 1738  
 QY 714 -MTTAGKALLSKNSURQEKQKFLQEAQFFLGRVLDKAIA-----KAEKAL---VTK 762  
 Db 1739 SIVNATKSLKEK-----QVSPEKTTQEDTPKENKDTSEVLAANKALKEKADKLSHLDTTK 1794  
 QY 763 KATKNGQOLLERSINKAVLAYNNSAIKKANVKELEKELDILGLVSGKGLAQAT-MVQGV 821  
 Db 1795 YTSSEANLSNALKKVNOVLTN---KDANKAQVQEAOLSL---QAEKDLVEKTESKDV 1848  
 QY 822 YLLKPLPLPEYIIGLN-----VYFDKSGKLIYALDMSPTIGEG---OKDAGNPIIL 870  
 Db 1849 ETAKQLREB-----LNKHQEDKQYTDSSAKV---KDAEKISEGVILDSKA----- 1894  
 QY 871 NVDENEGVALAVA-----TLADYEGLD----- 894  
 Db 1895 QADEVNAKADSLVEAEKGLVKEENKPAESDTEQVDVKAREALEQEVNKNANVNLIDGYTPE 1954  
 QY 895 -----IKTILNSKLSQTSIRQVPTAAVHRAGIFQAIONAA--AEAQLLPKP- 940  
 Db 1955 SODKPEILNGVRDVLNDKNASASLEKAEKVLETATGVLTVQVEHQVELPKVEQPVVTPE 2014  
 QY 941 -----GTHSEKSSSESANSKDRGLQSNPKTNRGHRSAILPRTGS 980  
 Db 2015 KQQTSEAEAKSESSESVSTSKDEVKEPEEK-KQDAHSVSDKGTGT 2058

## RESULT 72

AAU34339  
 ID AAU34339 standard; protein; 2434 AA.

XX AC AAU34339;

XX DT 14-FEB-2002 (first entry)

XX DE Staphylococcus aureus cellular proliferation protein #615.

XX KW Antisense; prokaryotic cellular proliferation protein; antibiotic;

XX KW antibacterial; drug design.

XX OS Staphylococcus aureus.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US009180.

XX PR 21-MAR-2000; 2000US-0191078P.

XX PR 23-MAY-2000; 2000US-0206848P.

XX PR 26-MAY-2000; 2000US-0207727P.

XX PR 23-OCT-2000; 2000US-0242578P.

XX PR 27-NOV-2000; 2000US-0253625P.

XX PR 22-DEC-2000; 2000US-0257931P.

XX PR 16-FEB-2001; 2001US-0269308P.

XX XX (ELIT-) ELITRA PHARM INC.

XX XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX XX WPI; 2001-611495/70.

XX DR N-PSDB; AAS2198.

XX PT New polynucleotides for the identification and development of

XX PT antibiotics, comprise sequences of antisense nucleic acids.

XX XX Example 3; SEQ ID NO 5835; 511pp; English.

XX XX The invention relates to antisense inhibitors of genes essential to

XX CC prokaryotic cellular proliferation, their use in identifying the genes,

XX CC their use in the discovery of novel antibiotics, the essential genes

CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
 CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence represents an essential prokaryotic  
 CC cellular proliferation protein. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 2434 AA;

Query Match 3.8%; Score 194; DB 4; Length 2434;

Best Local Similarity 19.5%; Pred. No. 0.019;

Matches 226; Conservative 177; Mismatches 454; Indels 300; Gaps 52;

QY 1 MKKELTKVALTITVSVVTHNQEVFSLVKEPILKQTOASSSI-SCADYAESGK----- 53

Db 226 LKEQVQS-ATTLDGVQTVKNSQTNTAMKGLRDSIANEATIKAGONYTDPASPNRNEYD 284

QY 54 -----SKLKINETSGPV--DDTVDFSDKRTTPEKIKDNLAKPRE-OELKAVNTES 105

Db 285 SAVTAAKAIINQTSNTPMTPEPTITQVTSQVTT-----KEQALNGARNLAQAKTTAKNNLN 339

QY 106 EKQITSGSQLESKESLSLNKTPSTNWEICDFITKGNLTVLGLSKSVEKLSQTDHL-- 163

Db 340 NLTSINNAKQDALTRSDIGATTAGV-NOETAKAKATELNNAHSL-QNGINDETQTKQTK 397

QY 164 ---VLPQADGTQLIQVASPAFTPK-----KTAIAEYTSRAGENGESIQLDVGKII 215

Db 398 YLDAEPSKKSAYDQAVNAAKAILTKASQNVDAV---EQALQNVNSTKTALNGDAKL 453

QY 216 NEGEVFNYSLLKKVTITPGYKH--GQDAFVDNKNIAEVNLPESLETISDYAPHAHLAKQ 273

Db 454 NEAKAAAKQTLTGLT-----HINNAQRTALDNE-ITQATNVEGVNTYKAKA-----QQ 500

QY 274 IDLPNLKAIGELAFFDNOITGK--LSLPQRLMLAERAFKSNHIKTIEFR-----G 323

Db 501 LD-----GAMQGL-----ETSIDKDTTLOSQNYQADDAKRTAYSQAVNAATILNKTAGG 552

QY 324 NSLKVIGEASFQ-----DNDLSQLMLPDGLEKIESEFTGNPGDDHNNRVVL---WTKS 375

Db 553 NTPKADVERAMQAVTQANTALNGIQNLDRAKQAANTAIT-NASDLNTKQKEALKAQVTS 611

QY 376 GRNPSGLATENT-----YNPDKSLWQESPEIDYTKWLE 409

Db 612 GRVSAANGVEHTATLTNTAMTALKRAIADKAETKASGNVYNADANKRQA-----YD 662

QY 410 EDFTYQKNSVTG-----FSNKGLO-----KVKRNKNLEIPKQH-----NGVT-- 446

Db 663 EKVTAENIVSGTPTPTLTTPADVTNAATQVTNAKTQLNGNHNLEVAQNANTADGLTSL 722

QY 447 -----ITEIGDNFRNVDVFQNTLRYDYLEEKLPS----- 477

Db 723 NGPQAKLKEQVGQATLTPNVQTVRDNA-QTLNTAMKGLRDSIANEATIKAGONYTDPASQ 781

QY 478 -----TIRK--IGAPAFQSNLKSFEASDLEIKEGAF---MNNRIETLELKOK 522

Db 782 NKQTDYNSAVTAAKAIIGQTTPSPMNAQEIINQAKQVTAQQAALNGOENLTATQNAKQH 841

QY 523 LVTIGD-----AAPHINIIYAIIVLSPVQEIGRSAPFRONGANNLIFMGSKV 568

Db 842 LNLGSLDLTAQDAVKAKQIEGATHNEV-----TQAQNNADALTAMTNL 886

QY 569 KT-LGEMAFILNRLHLDLSEKQLTRIPVQAFSDNALKEVLLPASLKTIREAPK----- 623

Db 887 KNGIQDQNTIKQGVNFTDADEAKRNATVNAVTQAEQILNKAQGPNTSKDGVETALENVQR 946



Db 1885 PVNRKASAREQLTTL-----FND-----KKQAIENAIQATVEE--RNSILAQLQ---NIYDT 1932  
 QY 640 AFNALDDNDGDQFN--KVVVKTTHNSVALADGSHFIVDPDKLSTIVDLEKILKLEGG 697  
 Db 1933 AIGQIDQDRSNAQVDKTASLNLQTHDL-----DVHPICKPDABKTINDDLARVTALVQ- 1986  
 QY 698 LDYSTLRQTOTQFQDMTTAGKALLSKSNLROGEKQKFLQEAQFFLGRVLDLKATAKAEK 757  
 Db 1987 -----NYRKVSNNRK-----ADALK 2001  
 QY 758 ALVTKKATKNGOLLERSIN--KAVLAYNNSAIKKNVKRLEKDLTLGLVEGKGLAQ 814  
 Db 2002 AITALKQWDEBELKTARTNADVDVLRFPNVALSDIEAVITEKNSLLR--IDN---IAQ 2056  
 QY 815 ATMVGCVLLKTPLEPPEYIGLVYFDKSGKLIYALOMSDTIGGQKDAYGNPILNVD 874  
 Db 2057 QT-----YAKFRAIATPEQLAKVKVLIDQ-----YVAD-----GNR--IDE 2091  
 QY 875 DNEGHALAVATLADYEGLDIKTLNLSQLTSIRQVPTAYHRAGIFQALQNAAAEAE 934  
 Db 2092 D-----ATLN-----DIKQHTQFVIDEILAIK-LPAEA-----TKVSPKEI 2126  
 QY 935 QLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG 982  
 Db 2127 QPAPKVCTPIKKEETHESRKVEK-----LPTGSEG 2158

## RESULT 74

AAR72826  
 ID AAR72826 standard; protein; 2482 AA.  
 XX  
 AC AAR72826;  
 XX  
 DT 27-FEB-1996 (first entry)  
 XX  
 DE Human mitosis.  
 XX  
 KW Cell cycle; M phase; mitosis; retinoblastome; mitosis; cell growth; inhibition.  
 KW  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT 1480..1659 Location/Qualifiers  
 FT /label= internal\_repeat  
 FT 1660..1839  
 FT /label= internal\_repeat  
 XX  
 PN W09511309-A2.  
 XX  
 PD 27-APR-1995.  
 XX  
 PF 24-OCT-1994; 94WO-US012162.  
 XX  
 PR 22-OCT-1993; 93US-00141239.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Zhu X, Lee W;  
 XX  
 DR WPI; 1995-170229/22.  
 DR N-PSDB; AAQ86851.  
 XX  
 PT Purified mammalian protein mitosis and agents that bind it and inhibit its action - used to promote cell growth or to inhibit cell division and/or proliferation.  
 PT  
 PS Claim 4; Fig 8B; 61pp; English.  
 XX  
 CC AAR72829 is human mitosis. Mitosis is involved in the regulation of the mammalian mitotic cell cycle. Mitosis is with 52F-1 (see AAR72824)  
 CC interacts with the retinoblastoma protein (the retinoblastoma tumour suppressor gene product). Mitosis is first synthesised at the G1/S

CC boundary, it is then phosphorylated from S through M phase, and during mitosis, is closely associated with the centromeres/kinetochore at the mitotic spindle poles. Mitosis is necessary for a eukaryotic cell to enter the M phase of the mitotic cell cycle and its degradation is necessary for a cell to advance on to the next stage. Mitosis is thus useful for controlling cell growth as overexpression of mitosis prevents a cell from exiting the M phase. An anti-mitosis antibody, antibody fragment or a phosphorylated mitosis mutagen ( or nucleic acid encoding it) can also be used to inhibit cell division which is particularly useful for the study of the cell cycle. A further use is to control hyperproliferative cells, and so control diseases such as psoriasis and breast cancer. It can also be used to block gametogenesis of an immature gamete  
 CC  
 XX  
 SQ Sequence 2482 AA;  
 Query Match 3.8%; Score 192.5; DB 2; Length 2482;  
 Best Local Similarity 20.2%; Pred. No. 0.025;  
 Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;  
 QY 6 KTVALTITTVSVVTHNQEVFSLVKEPIKQTAQSSISGADYAESGSKKIKINETSPV 65  
 Db 1341 KTTALDQLSEKMKKTKQEL-----ESHQSECLHCIQVAEAEVKEKTELLQT---L 1387  
 QY 66 DDTVTDLFSDKRTTPEKI---KDNLAGPREQELKAVTENTSEKQITSSQLEQSKES 121  
 Db 1388 SSDVSELLKDKTHLOEKLSLEKDSQALSUTKCELENIQALNKEKELL-----VKES 1440  
 QY 122 LSLNKTVPSTSNWEICD-----FITKGNLTVLGLS-----KSGVEKLS---QTD 161  
 Db 1441 ESLOARL-SESDYEKLNVSKALEAALVEKGEFALRLSTQBEVHQLRGIEKLAVRIAD 1499  
 QY 162 -----HLV--LPQAAADGTQLIQVASFAPTPDKKTAIABYTSRAGENGESQLDVG--- 211  
 Db 1500 EKKQLHTAEKLERENDSL-----KDKVENLERELQMSENGQELVILDAENSKA 1550  
 QY 212 -----KEIINEGEVFNYSLLKVTIPTGYKHI-----GQ-----DAF 243  
 Db 1551 EVETLKTQIEBMAEASLKVFE---LDLVTLRSEKNTLNQIEKQSQSELDKLSSFFKSL 1607  
 QY 244 VDNKNIAEVLNPLESLETISDYAFAPHAHLAKQIDLPNLKAIGELAFPNQITGKLSLRQL 303  
 Db 1608 LEEKEQAEIQIKESKTAVEM-----LQNLKEL-----NEAVALCGDQEI 1649  
 QY 304 MRLAERAPKSNHIKTIIFRGNLSKVIGEASQDNDLSQMLPDLGLEKIESAFTGNPDGD 363  
 Db 1650 MKATEQSLDPPPIEBEHQLR-NSIEKL-RARLEADEKQQLCVLQQLKESEHHA-----D 1700  
 QY 364 HYNRV-----VLWTKSGKNPSGLATENTVVPDKSLWQESPEIDYTKWLEEDFTYQKN 417  
 Db 1701 LLKGRVENLERELRIARTNQEHAALEAENS-----KGEVETLKAIEGTQ--- 1746  
 QY 418 SVTGFSNKGLO-----KVRNKNLEIPKQHNQVTTITEIGDNAPFRNVDFQNTLRKYDL 470  
 Db 1747 -----SLRGLELDVVTIRSEKENTNELQKEQERISELEIINSSFENI-LQEQEKVQM 1800  
 QY 471 BEVKLPSTIRKIGAFQFOSNNLSKFASDDLEETK-EGAFPMNRIETLEKDKLVTTGDA 529  
 Db 1801 KEKS--STAMEMLQTLKELNERNVAALHNDQEAQKAKQNLSSQVECELEKQALLQLD 1858  
 QY 530 APHINHIYIVLPESV-----QEIGRSAPRQNGANNLIFMGSKVKTGLGEMAFLSNRLEHD 585  
 Db 1859 EAKNY---IVLQSSVNGLIQEV-----EDGKQKL-----EKKDEISLKNQIQ--- 1900  
 QY 586 LSEQQLKTEIPVQAFSDNAL-KEVLLPASLKTITREAFKKNHLKQLEVASALSIAFNAL 644  
 Db 1901 --DQEQVLVSKLSQVEGEHQLWKEQNL--ELANLTVLEQKIQVLQSKNAS-----L 1947  
 QY 645 DDNDGDEQDFDNKVVVTKTHNSYALADGEHFIVDPDKLSTIVDLEKILKLEGLDYSLR 704  
 Db 1948 QD-----TLEVLQSSYKKNLENELELTQDKMSF-----VEKVNKM-----TAK 1985  
 QY 705 QT-TQTQFRDMMTAGKALLSKSNLROGEKQKFLQEAQFFLGRVLDLKATAKALVTKK 763

Db 1986 ETELOREHEM--AQKTAELQEBL-SGEKNRLAGELQLLEI-----K 2026  
Qy 764 ATKNGOLLERSINKAVLAYNNSAIKKA---NVKLEKELDLDATGLVEGKGPLAQATMVQ 819  
Db 2027 SSXD-QKE-----LTENSELKSLDCMHKQOQVEGKVRRETAETOLRHEAEKXH 2078  
Qy 820 GVYLLKTPPLPEYIYGLNVYFDK-----SGKLIYALDMSDTIG 858  
Db 2079 QALLLDTN---KQYEVEIOTYREKLTSSKECLSSQKLEIDLKSSKEELNNSLKATTOIL 2135  
Qy 859 EGOXDA-----YGNPILNVDENEGHALAVATLADYEGLDIKTILNSKLSOL----- 906  
Db 2136 ELUKTKMDNLKYVNLQKXENERAQGMKLLIKSCKQLE--BEKEILOKELSOLAQOEK 2193  
Qy 907 -----TSIROVPTAAVHRAGIFQIONAAAEAPQLPKPG-----THSEKSSSESAN 954  
Db 2194 QKTGTWMDYKVBELTTEIKE---LKTLEKTKTEADYDLKYCSLLISHKELEKAKEMLE 2250  
Qy 955 SKDRGLQSNPKTNRGHRSAIL 975  
Db 2251 TQVAHLCSQOSKQDSRGSPLL 2271

RESULT 75  
AAW23396  
ID AAW23996 standard; protein; 2482 AA.

AC AAW23996;

XX 28-MAY-1998 (first entry)

DE Human mitosis amino acid sequence.

KW Mitosis; phosphoprotein; mitotic cell cycle; antibody; analogue;  
KW inhibition; M phase; Antagonist; hyperproliferative cell; cancer;  
KW leukaemia; lymphoma; Chromosome segregation.

XX Homo sapiens.

PH Key Location/Qualifiers

FT Domain 258..280  
/note= "leucine heptad repeat"

FT Domain 340..362

FT Domain 564..593

FT Domain 1387..1443

FT Domain 1885..1962

FT Domain 2146..2188

FT Domain 2165..2187

FT Misc-difference 2188  
/note= "leucine heptad repeat"

FT Misc-difference 2189

FT Misc-difference 2300

FT /label= Bipartite targeting motif

FT /note= "Optionally C or G"

FT Misc-difference 2301

FT Misc-difference 2303

FT /label= Bipartite targeting motif

FT /note= "Optionally A or T"

XX US5710022-A.

PN 20-JAN-1998.

PD 24-OCT-1994; 94US-00328254.

XX 22-OCT-1993; 93US-00141239.

XX (TEXA ) UNIV TEXAS SYSTEM.

PA Lee W, Zhu X;

XX WPI; 1998-109817/10.

XX

XX

XX

XX

XX

XX

XX

XX

DR N-PSDB; AAV09076.  
XX New isolated mitosis protein and gene - useful for, e.g. developing  
PT products for therapy and diagnosis of hyper-proliferative disorders such  
PT as cancers or psoriasis.  
XX Claim 1; Col 40-52; 43pp; English.  
XX This is the amino acid sequence for mitosis, a phosphoprotein necessary  
CC for the cell to enter mitosis. The protein's degradation is also necessary  
CC for the cell to advance into the next stages of mitosis. The mitosis  
CC protein, can be used to control the growth of cells. An anti-mitosis  
CC antibody, a mutant or a non-functional analogue of mitosis can inhibit  
CC the mitotic cell cycle by preventing the cells from entering the M phase,  
CC and over expression of mitosis or its functional equivalent, would  
CC inhibit the cycle by preventing cells from leaving the M phase.  
CC Antagonists to this protein can be used to control hyperproliferative  
CC cells in, (e.g. thyroid hyperplasia, Grave's disease, psoriasis, benign  
CC prostatic hypertrophy, Li-Fraumeni syndrome, breast cancer, sarcomas and  
CC other neoplasms, bladder cancer, colon cancer, lung cancer and various  
CC leukemias and lymphomas). Reintroduction or supplementation of lost  
CC mitosis function by introduction of the protein or nucleic acid encoding  
CC the protein into a cell can restore defective chromosome segregation,  
CC which is a marker of progressing malignancy. Malignant proliferation of  
CC cells can then be halted. The protein can also be used for the detection  
CC and diagnosis of hyperproliferative cells  
XX  
XX Sequence 2482 AA;

Query Match 3.8%; Score 192.5; DB 2; Length 2482;  
Best Local Similarity 20.2%; Pred. No. 0.025;  
Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;

Qy 6 KTVALTITVSVVTHNQEVFSLVKEPILKQTOASSISGADYAEAGSGSKLKNITSGPV 65  
Db 1341 KTTALDQLSEKMKETQEL-----ESHQSECLHCIOVAEAEVKEKTELLQT---L 1387  
Qy 66 DDTVTDLFSDKRTTPEKI---KDNLAGPPEQELKAVTENTSEKQITSGSQLEQSKS 121  
Db 1388 SSDVSELLKDKTHLQEKLSQLEKDSQALSLTCELENQIAQLNKEKEL-----VKES 1440  
Qy 122 LSLNKTVPTSTSNWEICD-----FITKNTLVGLS-----KSGVEKLS---QTD 161  
Db 1441 ESLQARL--SESDYKILNVSKALEAALVEKGEPALRLSTQEEVHQLRRGIEKLVRV 1499  
Qy 162 ----HLV--LPSQADGTQLIOVASFAFTPPKKTAAIAEYTSRAGENGISOLDVDG--- 211  
Db 1500 EKKQLHIAFKLERERENDSL-----KDKVENLERELQMSSEENQELVILDAENSKA 1550  
Qy 212 -----KEIINGEVFNSYLLKKVTIPTGYKHI-----GQ-----DAP 243  
Db 1551 EVETLKTQIEEAWRSKLVFE---LDLVLRSEKENLTQIQEKQQLSGLDLSFKSL 1607  
Qy 244 VDNKNIAEVLNPLETISDYAFALHAKLQIDLPNLKAIGELAPFDNQITGKLSLPRL 303  
Db 1608 LEEKQAEIQIKESKTAVEM-----LQNLKEL-----NEAVALCGDQEI 1649  
Qy 304 MRLAERAFKSNHIKTIIEFRGNSLKVIGEASFOODLSQLMLPDGLEKTESEFTNGPGD 363  
Db 1650 MKATEQSLDPPTIEEHLRL--NSIEKL--RARLEADEKKQLCVLQQLKESEHHA-----D 1700  
Qy 364 HYNNEV-----VLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKN 417  
Db 1701 LLKGRVENLERELEIARTNOEHALEAENS-----KGEVETLKAKEGTQ--- 1746  
Qy 418 SVTGFSNKGLO-----KVKENKNLEIPKQHNGVITIEIGNVAFRNVDFQNKTLRKYDL 470  
Db 1747 ----SLRGLELDVVTIRSEKENLTNELOKEQERISELEIINNSFENI--LOEKEQEKVQM 1800  
Qy 471 BEVKLPSTIRKIGAFAPFNSNNLSKFEASDDLEBEIK--EGAFMNNRIETLELKKLVTVIGA 529  
Db 1801 KEKS--STAMWMLQTLQKELNERNVAALHNDQACKAKENQNLSSQVECLELEKALQLOGID 1858





Db 2432 KEKS--STAMEMLQTLKELNERNVAALHNDQACKAKQNLSSQVECLELEKAQLQGLD 2489

Qy 530 AFHINHIYAIVLPESV---QBIGSAFRQNGANNLI FMGSKVKTIGEMAFISNRLHLD 585

Db 2490 EAKNNY---IVLQSSVNGLIQEV-----EDGKQL-----EKKDEISRLKNQIQ--- 2531

Qy 586 LSEKQLTPIPVQAFSDNAL-KEVLLPASLKTIRIEAFKKNHLKQLEVASALSHIAFNAL 644

Db 2532 --DQELVSKLSQVEGEHLWKEQNL--ELRNLTVELEQKIQVLOSKNAS-----L 2578

Qy 645 DNDGDGEQDNKVVVKTTHNSYALADGEHFI VDPDKLSSTIVDLKILKLEGLDYSTLR 704

Db 2579 QD-----TLEVLOSSYKNUENELELTMDKMSF-----VEKVNKM-----TAK 2616

Qy 705 QT-TQTFQDMTITAGKALLSKNSLRQEQKQFLQEAQFFLGRVDLDKAKAEKALVTKK 763

Db 2617 ETELQREMHM--AQKTAELOEEL-SGKNRLAGELQLLLEI-----K 2657

Qy 764 ATKNGQLLERSINKAVLAYNNSAIKKA-----NVKREKELDLTLGLVEGKGPLAQATWVO 819

Db 2658 SSKD-QLKE-----LTLENSELKSLDCMHKQDQVEKEGKVRERIEARYQLRHEAEKKH 2709

Qy 820 GYLLKTPPLPEYVIGLVNVPDK-----SGKLIYALDMSDTIG 858

Db 2710 QALLDNTN---KQYEVEIQTYREKLTSSKECLSSQKLEIDLKSSKBELNLSKATQTL 2766

Qy 859 EQGKDA-----YGNPILNVDEBEGYHALAVATLADYEGLDIKTILNSKLSQL----- 906

Db 2767 BELKKTMDNLKYVNLKKNERAGQKMKLLIKSCKQLE--BEKELQKELSOLQAQEK 2824

Qy 907 -----TSIRQVPTAAVHRAGIFQAIQNAAAEAEQLLPKP-----THSEKSSSESAN 954

Db 2825 QKTGTVMDTKVDELATTEIKE---LKETLEKTEADEYLDKYCSLLISHEKLEKAKEMLE 2881

Qy 955 SKDRGLQSNPKTNRGHSAIL 975

Db 2882 TQVAHLCSQSQKDSRGSPLL 2902

RESULT 77

ID ADK70220 standard; protein; 3113 AA.

XX AC ADK70220;

XX DT 06-MAY-2004 (first entry)

XX DE Human oesophageal cancer antigen SEQ ID NO:16.

XX KW human; oesophageal cancer; antigen.

XX OS Homo sapiens.

XX PN JP2003259872-A.

XX PD 16-SEP-2003.

XX PF 07-MAR-2002; 2002JP-00061793.

XX PR 07-MAR-2002; 2002JP-00061793.

XX PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX DR WPI: 2004-093976/10.

XX DR N-PSDB; ADK70219.

XX PT Novel human oesophageal cancer antigen peptide, useful for diagnosing

XX PT oesophageal cancer or oesophageal cancer high-risk person.

XX PS Example 3; SEQ ID NO 16; 70pp; Japanese.

XX CC The invention relates to a novel human oesophageal cancer antigen peptide

CC and the polynucleotide encoding it. A peptide of the invention is useful

CC for diagnosing oesophageal cancer or an oesophageal cancer high-risk

CC person. The method of diagnosis disclosed in the invention provides

CC effective diagnosis of oesophageal cancer. The present sequence

CC represents a human oesophageal cancer antigen peptide of the invention.

XX Sequence 3113 AA;

Qy Query Match 3.8%; Score 192.5; DB 8; Length 3113;

Best Local Similarity 20.2%; Pred. No. 0.034;

Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;

Qy 6 KTVALTTLTVSVVTHNQEVFSLVKEPILKQTOASSISGADYAESGSKLKINTSGV 65

Db 1972 KTTALDQLSEKMKETQEL-----ESHQSECLHCQVAAEVKEKTELLQT---L 2018

Qy 66 DDTVTDLSDKRTTPEKI---KDNLAGPREQELKAVTENTSEKQITSGSOLQSKEK 121

Db 2019 SSDVSELKDKTHLQELKQSLSEKDSQALSITKCELENOIAQNKKEKELL-----VKRS 2071

Qy 122 LSLNKTVPSTSNWEICD-----FITKGNLTVLGLS-----KSGVEKLS---QTD 161

Db 2072 ESLQARL--SESDYEKLVNSKALEAALVEKGFALRLSSTQBEVHQLRRGIEKLRVRIEAD 2130

Qy 162 -----HLV--LPSQAADGTQLIQVASFAPTPDKKTAIAEYTSRAGENGEISOLDVDG--- 211

Db 2131 EKKQLHIAEKLKERENDSL-----KDKVENLERELQMSSENOELVILDAENSKA 2181

Qy 212 -----KEIINEGEVFNLSYLLAKKVTIPITYKHI-----CQ-----DAF 243

Db 2182 EVETLTKTOIEMARSLKVFE---LTLVTRSEKENLTQIOEKQSELDELKLSFSKSL 2238

Qy 244 VDNKNIAEVLNLPESLETISDYAFALHALKQIDLPNLKAIGELAFDNOITCKLSLPROL 303

Db 2239 LEEKEQAIQIKESKTAVEM-----LQNLKEL-----NEAVALCGDQEI 2280

Qy 304 MRLAEAFKSNHKTIEFRGNSLKVIGEASFQDNDLSQLMPLDGLKEKIESEFTNGPDD 363

Db 2281 MKATEQSLDPPTEEHQRLR-NSIEKL--RARLEADEKKQLCVLQQLKESEHHA-----D 2331

Qy 364 HYNRRV-----VLWTKSGKNPGLATENTVYVNPDKLSLWQESPEIDTYKWLDEEDTYQKN 417

Db 2332 LLKGRVENLERELEIARTNQEHAALEAENS-----KGEVETLKAKTEGTMQ--- 2377

Qy 418 SVTGFSGKGLQ-----KVKRNKNLEIPKHNGVTITTEIGDNAPFNVDFOKTLKYDL 470

Db 2378 -----SLAGLELDVVTIRSEKENLTNLOKQERISELEIINSSPENI--LQEKQEKVQM 2431

Qy 471 EEVKLPSTIRKIGAFAPQSNLKSFEASDDLEEIK-EGAPMNNRIETLEKDKLVITGDA 529

Db 2432 KEKS--STAMEMLQTLKELNERNVAALHNDQACKAKQNLSSQVECLELEKAQLQGLD 2489

Qy 530 AFHINHIYAIVLPESV---QBIGSAFRQNGANNLI FMGSKVKTIGEMAFISNRLHLD 585

Db 2490 EAKNNY---IVLQSSVNGLIQEV-----EDGKQL-----EKKDEISRLKNQIQ--- 2531

Qy 586 LSEKQLTPIPVQAFSDNAL-KEVLLPASLKTIRIEAFKKNHLKQLEVASALSHIAFNAL 644

Db 2532 --DQELVSKLSQVEGEHLWKEQNL--ELRNLTVELEQKIQVLOSKNAS-----L 2578

Qy 645 DNDGDGEQDNKVVVKTTHNSYALADGEHFI VDPDKLSSTIVDLKILKLEGLDYSTLR 704

Db 2579 QD-----TLEVLOSSYKNUENELELTMDKMSF-----VEKVNKM-----TAK 2616

Qy 705 QT-TQTFQDMTITAGKALLSKNSLRQEQKQFLQEAQFFLGRVDLDKAKAEKALVTKK 763

Db 2617 ETELQREMHM--AQKTAELOEEL-SGKNRLAGELQLLLEI-----K 2657

Qy 764 ATKNGQLLERSINKAVLAYNNSAIKKA-----NVKREKELDLTLGLVEGKGPLAQATWVO 819

Db 2658 SSKD-QLKE-----LTLENSELKSLDCMHKQDQVEKEGKVRERIEARYQLRHEAEKKH 2709

Qy 820 GYLLKTPPLPEYVIGLVNVPDK-----SGKLIYALDMSDTIG 858

Db 2710 QALLDNTN---KQYVEIQTREKLTSEKCLSSQKLEIDLLKSSKEBNNLSKATTOIL 2766

Qy 859 EGQKDA-----YGNPIINVEDNRYHALAVATLADYEGLDIKTILNSKLSQL----- 906

Db 2767 BELKTKMDNLKYVNLQKENERAQCKMLIKSCCKQLE--EKEILQKLSQLQAAQEK 2824

Qy 907 -----TSIRQVPTAAYHRAGIFQAIQNAABAEQQLPKPG-----THSEKSSSESAN 954

Db 2825 QKTGVMTQKVDLTTTEIKE---LKETLEKTKADEYLDKYCSLLISHEKLEKAKEMLE 2881

Qy 955 SKDRGLQSNPKTNRGHRSAIL 975

Db 2882 TQVAHLCSQSQKDSRGSPL 2902

RESULT 78

ID ADL72172

XX ADL72172 standard; protein; 3113 AA.

XX AC ADL72172;

XX DT 20-MAY-2004 (first entry)

XX Human solid-cancer antigen peptide mitosis.

XX DE Solid-cancer antigen; FIR; CENP-A; HOOK2; myomegalin; MKRN1; KIAA1545;

XX KW enigma; TROP2; mitosis; CU-EC-1; cytostatic; solid cancer; human.

XX OS Homo sapiens.

XX PN WO2004018518-A1.

XX PD 04-MAR-2004.

XX PF 21-APR-2003; 2003WO-JP005046.

XX PR 23-AUG-2002; 2002JP-00244249.

XX PA (NTSC-) JAPAN SCI & TECHNOLOGY CORP.

XX PI Shimada H, Hiwasa T, Tomonaga T, Matsushita K, Nomura F;

XX PI Takiguchi M, Ochiai T;

XX DR WPI; 2004-248065/23.

XX DR N-PSDB; ADL72171.

XX DR GENBANK; U30872.

XX Human solid cancer antigen peptides and polynucleotides useful in early diagnosis and treating solid cancer.

XX PT Disclosure; SEQ ID NO 18; 136pp; Japanese.

XX PS The invention relates to human solid-cancer antigen peptides and encoding polynucleotides. The antigen peptides are selected from FIR, CENP-A, HOOK2, myomegalin, MKRN1, KIAA1545, enigma, TROP2, mitosis, CU-EC-1, Cytostatic, the antigen peptides and their polynucleotides are applicable in early diagnosis and useful in treating solid cancer. With these 10 antigen peptides and their encoded polynucleotides, early diagnosis of solid cancer can be achieved accurately and quickly. The present sequence represents a human solid-cancer antigen peptide, mitosis.

XX SQ Sequence 3113 AA;

Query Match 3.8%; Score 192.5; DB 8; Length 3113;

Best Local Similarity 20.2%; Pred. No. 0.034;

Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;

Qy 6 KTVALTLTVSVTNQEVSLVKEPILKQTQASSISGADYAESGSGKSKINETSQPV 65

Db 1972 KTTALDQLSEKMKERTQSL-----ESHQSECLHCIQVAEAEVKEKTELLQT---L 2018

Qy 66 DDTVDTLPSDKRTTPEKI-----KDLNKGPRQELKAVNTSESEKQITSGSQLESKES 121

Db 2019 SSDVSELLKDKTHLOEKLOSLEKDSQALSILTKCELENQIAQLNKEKELL-----VKES 2071

Qy 122 LSLNKTVPSTSNWICD-----FITKGNLTVLGS-----KSGVEKLS---QTD 161

Db 2072 ESLQARL-SESDEKLVNSKALEAALVEKGEFALRLSSTQEBVHQLRGIEKLRVREAD 2130

Qy 162 -----HLV--LPSQAADGTQLIQVASFAFTPDKKTATAEYTSRAGENGESQLDVG--- 211

Db 2131 EKKQLHTAELKRERENDSL-----KDKVENLERELQWSEENQOELVILDAENSKA 2181

Qy 212 -----KEIINEGEVFNYSYLLKVTIPTGYKHI-----GQ-----DAF 243

Db 2182 EVETLKTQIEBARSLLKVF-----LDIVTLRSEKENLTQIQEKGQGLSELDKLSSPFSK 2238

Qy 244 VDNKNIABVNLPSLETISDYAFAHALAKQIDLDPNLIKAGELAFPNQITGKLSLPRQL 303

Db 2239 LEEKEQAEIQIKESKTAVEM-----LQNLKEL-----NEAVALCGDQEI 2280

Qy 304 MRLAERAPKSNHIKTIEFRGNSLKVIGEASFQDNDLSQLMPLDGLKIESEAFNGPDOD 363

Db 2281 MKATEQSLDPPIEEHQLR-NSIEKL-RARLEADEKKQLCVLQQLKSEHHA-----D 2331

Qy 364 HYNRV-----VLWTKSGKNPSGLATENTVVPDKSLWQESPEIDYTKMLEEFTYQKN 417

Db 2332 LLKGRVENLERELEIARTNQEHAALEANS-----KGEVETLKAKIEGTQ--- 2377

Qy 418 SVTGFSNKGLO-----KVRNKNLEIPKOHNGVTTITEIGDNAPRVDFQNTLRKYDL 470

Db 2378 -----SLRGLELDVVTIRSEKENLTNELQKEQERISELEIINSSFENI-LQEKEQKQV 2431

Qy 471 BEVKLPSTIRKIGAPAFQSNLKSFEASDDLEIK-EGAPMNNRIETLELKDLVLTGDA 529

Db 2432 KEKS--STAMEMLQTLKELNVAALHNDQEAQKAKEQNLSSQVECLELEKALQLOGLD 2489

Qy 530 APHINHIVAIPLPESV-----QEIGRSAPFRONGANNLIFMGSKVKTGLGMAFLSNRLEHD 585

Db 2490 EAKNY-----IVLQSSVNGLIQEV-----EDGKQKL-----EKKDEISLKNQIQ--- 2531

Qy 586 LSEKQLTEIPVQAFSDNAL-KEVLLPASLKTIRREAFKKNHLKOLEVASALSIAFNAL 644

Db 2532 --DQEQLVSKLSQVEGEBHQLWKEQN--ELRNLTVLEQKIQVLQSKNAS-----L 2578

Qy 645 DDNDGDSQFQKVVVTKHNSYALADGEHFTVDPDKLSSSTIVDLEKILKLEGLDYSTLR 704

Db 2579 QD-----TLEVLSQSYKNELELELTMDKMSF-----VEKYNKM-----TAK 2616

Qy 705 QT-TOTQFRDMTTAGKALLSKSNLROGEKOKFLQEAQFFLGRVLDLDAIAKAEKALVTKK 763

Db 2617 ETELQREHHEM--AQKTAELQEEEL-SGEKNRLAGELQLLEEI-----K 2657

Qy 764 ATKNGOLLERSINKAVLAYNNSAIKKA-----NVKRLKEKLDLLTGLVEGKGPLAQATMVQ 819

Db 2658 SSKD-QLKE-----LTENSELKSLDCHMKHQVEKGVREBIEAYQLRHEAEKKH 2709

Qy 820 GVIYLLKTPPLPYEYIGLVNVPDK-----SGKLIYALDMSDTIG 858

Db 2710 QALLDNTN---KQYVEIQTREKLTSEKCLSQKLEIDLLKSSKEBNNLSKATTOIL 2766

Qy 859 EGQKDA-----YGNPIINVEDNRYHALAVATLADYEGLDIKTILNSKLSQL----- 906

Db 2767 BELKTKMDNLKYVNLQKENERAQCKMLIKSCCKQLE--EKEILQKLSQLQAAQEK 2824

Qy 907 -----TSIRQVPTAAYHRAGIFQAIQNAABAEQQLPKPG-----THSEKSSSESAN 954

Db 2825 QKTGVMTQKVDLTTTEIKE---LKETLEKTKADEYLDKYCSLLISHEKLEKAKEMLE 2881

Qy 955 SKDRGLQSNPKTNRGHRSAIL 975

Db 2882 TQVAHLCSQSQKDSRGSPL 2902

RESULT 79

ADQ18045

ID ADQ18045 standard; protein; 3113 AA.  
 XX AC ADQ18045;  
 XX DT 26-AUG-2004 (first entry)  
 XX DE Human soft tissue sarcoma-upregulated protein - SEQ ID 862.  
 XX KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.  
 XX OS Homo sapiens.  
 XX PN WO2004048938-A2.  
 XX PD 10-JUN-2004.  
 XX PF 26-NOV-2003; 2003WO-US038193.  
 XX PR 26-NOV-2002; 2002US-0429739P.  
 XX PA (PROT-) PROTEIN DESIGN LABS INC.  
 XX PI Aziz N, Ginsburg WM, Zlotnik A;  
 XX WIPI; 2004-441208/41.  
 XX DT Early detection of soft tissue sarcoma comprises determining expression  
 PT of a gene in a first soft tissue sample and a normal soft tissue sample  
 PT and comparing the gene expression, also useful in treating soft tissue  
 PT sarcoma.  
 XX PS Example 2; SEQ ID NO 862; 210pp; English.  
 XX CC The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples and comparing the  
 CC expression of the gene in both soft tissue samples, where a higher level  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cytostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC protein of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.  
 XX Sequence 3113 AA;  
 SQ

Query Match 3.8%; Score 192.5; DB 8; Length 3113;  
 Best Local Similarity 20.2%; Pred. No. 0.034;  
 Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;

QY 6 KTVALTTVTVVTHNQVFLVKEPIKQTAQSSISGADYAEVSGSKJINETSQPV 65  
 DB 1972 KTTALDQLSEKMKKEKTEQL-----ESHQSECLHICIQVAEAEVKEKTELLQT---L 2018

QY 66 DDTVTDLPSDKETTPK-----KONLAKGPRQELKAVNTESKEQITSGSQLESKES 121  
 DB 2019 SBDVSELLKDKTHQELQKLSQKDSQALSITKCELENQAOLNKEKELL-----VKES 2071

QY 122 LSLNKTVPSTSNWEICD-----FITGNTLVGLS-----KSGVEKLS---QTD 161  
 DB 2072 ESLQARL-SESDYEKLNVSKALEALVEKGEFALRLSSQTEVHQVLRRIEKLVRVIEAD 2130

QY 162 -----HLV--LPSQADGTQLIQVASFAPTPDKTAIEYTSRAGNGEISQLDVDG---- 211  
 DB 2131 EKKQLHIAKLEKLERENDSL-----KDKVENLERELQWSENQELVILDAENSKA 2181

QY 212 -----KEINEGEVNSYLLKKVTPTGKHI-----GQ-----DAF 243  
 DB 2182 EVETLKTQIEEMARSLKVPE---LQVLTRSEKENLTKQIQEKQGLSELDKLLSFKSL 2238

QY 244 VDNKNIAEVNLPESLETISDYAFAPHALKQIDLPDNLKAIGELAFDDNQITGKLSIPROL 303  
 DB 2239 LEEKEQAEIQIKESKTAVEM-----LQNLKEL-----NEAVALCGDQRI 2280

QY 304 MRLAERAFKSNHIKTIIEFRGNSLKVIGASFOVDNLSQLMLPDGLEKIESEAFNGPDD 363  
 DB 2281 MKATEQSILDPPIIEEHQLR-NSIEKL-RARLEADEKKQLCVLQQLKESHHH-----D 2331

QY 364 HYNRV-----VLMTGSGKPSGLATENTYVNPDKSLWOESPEIDYTKWLSEEDFTYQKN 417  
 DB 2332 LLKGRVENLERELEIARTNQEHAALEANS-----KGEVETLKAKIEGTMQ--- 2377

QY 418 SVTGFSNKGLO-----KVRKNKNLEIPKQHNGVTITIEGDNAFRNVDQFNKTLRKYDL 470  
 DB 2378 -----SLRGLELDVVVTIRSEKENLTNELQKEQERISELEIINSFENI-LQKEQEKVQM 2431

QY 471 BEVKLPSTIRKIGAFAPFOSNNLKSFEASDDLEIK-EGAFPMNRITETLEKDKLVITGDA 529  
 DB 2432 KEKS--STAMENLOTQLKELNERNVAALHNDQACAKQENLSSQVECELEKALQLQGLD 2489

QY 530 APHINHIAIVLPESV---OEIGRSAFRONGANNLIIFMGSKVKTLGEMAFPSNRLEHLD 585  
 DB 2490 EAKNY-----IVLQSSVNGLIQEV-----EDGKQKL-----EKDEEISRLKNQIQ--- 2531

QY 586 LSEKQLTEIPVQAFSDNAL-KEVLLPASLKTIRREEAFKQHLKQLEVASALSHIAFNAL 644  
 DB 2532 --DQQLVSKLSQVEGHEQLMKEQNL--ELRNLTVLEQKIQLVQSKNAS-----L 2578

QY 645 DDNDGDEQDNKVVVTHNSYALADGHEFIVDPKLSSTIVDLKILKLEGLDYSTUR 704  
 DB 2579 QD-----TLEVLQSSYKNLENELELTMDKMSF---VEKVNKM-----TAK 2616

QY 705 QT-TOTQPRDMTTACKALLSKSNLRQGEKQKFLQEAQFPLGRVLDLKAIAEAKALVTYK 763  
 DB 2617 ETELOREHEM--AOKTAELOEEL-SGEKNRLAGELQLLLEEI-----K 2657

QY 764 ATKNQQLLERSINKAVLAYNNSAIKKA---NVKLEKELDLTLGLVGEKGPLAQATVMQ 819  
 DB 2658 SSKD-QLAKS-----LTLENSELKSLDCMHKQDQVEKEGKVRBEIAEVQLRHEAKGH 2709

QY 820 GYLLKTPPLPPEYIIGLVNVPDK-----SGKLIYALDMSDTIG 858  
 DB 2710 QALLDNTN---KQYEVEIQTREKLTTSKECLSSQKLEIDLKSKKEELNNSLKATQTIL 2766

QY 859 EGQKDA-----YGNPILNVDEDNQYHALAVATLADYEGLDIKTILNSKLSQ----- 906  
 DB 2767 BELKTKMDNLKYNQLKKNERAGQKMKLLIKSCQLE--EKEILQKELSQLQAQEK 2824

QY 907 -----TSIRQVPTAAYHRAGIFQAIQNAAAAEQQLPKPG-----THSEKSSSESAN 954  
 DB 2825 QKTGTVMDTKVDELTEIKE---LKETLEEKTEADEVIDKYCSLLISHKLEKAKEMLE 2881

QY 955 SKDRGLQNPKNTRGRHSAIL 975  
 DB 2882 TQVAHLCSQQSKQDSRGSPLL 2902

RESULT 80  
 AAR99795  
 ID AAR99795 standard; protein; 3248 AA.  
 XX AC AAR99795;  
 XX DT 08-OCT-1996 (first entry)  
 XX DE Kinetochore protein CENP-F.  
 XX KW Kinetochore protein; CENP-F; cell cycle; cancer; diagnosis;  
 XX KW autoimmune antibody.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers

FT Domain 1. .200  
 FT /label= Extended\_coiled\_structure  
 FT 280. .1350  
 FT /label= Extended\_coiled\_structure  
 FT 1380. .1610  
 FT /label= Globular\_domain  
 FT /note= "globular\_domain consists of 2 direct repeats of  
 FT 95 amino acids"  
 FT 1620. .1750  
 FT /label= Extended\_coiled\_structure  
 FT 1850. .2990  
 FT /label= Extended\_coiled\_structure  
 FT 3048. .3248  
 FT /label= C-terminal domain  
 FT /note= "the C-terminal domain is predicted to form a  
 FT proline-rich (10.6%) highly basic (pI 10) globular  
 FT domain"  
 FT XX  
 PN W09617867-A1.  
 XX  
 XX 13-JUN-1996.  
 PD  
 XX 08-DEC-1995; 95WO-US016216.  
 PF  
 XX 09-DEC-1994; 94US-00353700.  
 PR  
 XX (FOX-C) FOX CHASE CANCER CENT.  
 PA (UYTE-) UNIV TECHNOLOGIES INT INC.  
 XX  
 PI Yen TJ, Rattner JB;  
 XX  
 XX WPI: 1996-287116/29.  
 DR N-PSDB; AAT34578.  
 XX  
 PT DNA encoding kinetochore protein - used as a marker for the G2 and M  
 PT phases of a cell cycle, partic. for detection of malignant diseases.  
 PT  
 XX Claim 12; Page 41-54; 72pp; English.  
 PS  
 XX A 372 kDa human kinetochore protein, CENP-F (AAR99795), is detected by  
 CC immunofluorescence microscopy only during the G2 and M phases of a cell  
 CC cycle. It is the product of a cDNA clone (AAT34578) isolated from a  
 CC breast carcinoma cDNA library. Recombinant CENP-F can be produced by  
 CC expression in prokaryotic or eukaryotic host cells. CENP-F can be used to  
 CC detect autoimmune antibodies to the protein, which may provide an early  
 CC diagnosis for the onset of various malignant diseases. Use of CENP-F as a  
 CC cell cycle marker allows the specific detection of G2 and M phase cells  
 XX  
 SQ Sequence 3248 AA;  
 Query Match 3.8%; Score 191.5; DB 2; Length 3248;  
 Best Local Similarity 20.1%; Pred. No. 0.041;  
 Matches 221; Conservative 179; Mismatches 400; Indels 301; Gaps 52;  
 QY 6 KTVALLTVSVVTHNQEVSLVKEPILKQTOASSISGADVAESSGSKKLINETSQPV 65  
 DB 2069 KTTALDQSEKKEKTKQEL-----ESHQSECLHCICQVAEAVKEKTELLQ---L 2115  
 QY 66 DDTVTLDFSDKRTTPEKI-----KDLAKGPREQELKAVTENTSEBKQITSGSLQESKES 121  
 DB 2116 SSDVSSELLKDKTHLQKLSLEKDSQALSITKCELENQIQLNKEKELL-----VKES 2168  
 QY 122 LSLNKTVPSTSNWEICD-----FTYKNTLVGLS-----KSGVEKLS---QTD 161  
 DB 2169 ESLQARL-SESDYKELNVS KALEALVEKGEFALRLSSTQBEVHQLRRGIEKLRVRIEAD 2227  
 QY 162 -----HLV--LPSQADGTLQIQVASFAPTPDKTAIAEYTSRAGENGESOLDVDG--- 211  
 DB 2228 EKKQLHIAEKLKERENDSL-----KDKVENLERELQMSENQELVILDENSKA 2278  
 QY 212 -----KEINEGEVFNYSLLKKVTIPTGYKHI-----QO-----DAF 243  
 DB 2279 EVETLTKQIEEMARSLKIFE---LDLVLRSEKENLTQIQEKQQLSELDKLLSFFKSL 2335

QY 244 VDNKNIAEVLNIPESLETISDYAFAPHALAKQIDIDPDNLKATGELAFPDNQITGKLSLPQL 303  
 DB 2336 LEEKEQABIQIKESKTAVEM-----LQNLKEL-----NEAVALCGDQEI 2377  
 QY 304 MRLAERAFKSNHITITIFRGNLSKVIGEASPDNDLSQMLPDGLEKIESAFTGNPGDD 363  
 DB 2378 MKATEQSLDPPIEEHQLR-NSIEKL-RARLEADEKKQLCVLQQLKSESHA-----D 2428  
 QY 364 HYNRRV-----VLWTKSGKNPSGLATENTYVVPDKSLMOBSPEDIDYTKMLBEEFTYQKN 417  
 DB 2429 LLKGRVENLERELEIARTNQSHAALAEANS-----KGBVETLKAKIEGTQ--- 2474  
 QY 418 SVTGFNSKGLQ-----KVKGNKNLRIPKOHNGVTITIGDNAPRVDFQNKTLRKVDL 470  
 DB 2475 -----SURGLELDVVTIRSEKENLTNLOKEQERISELEIINSSFENI-LOEKBOEKVQM 2528  
 QY 471 BEVKLPSTIRKIGAPAFQSNLKSFEASDDLEETK-EGAPMNNRIETLEKDKLVITIGDA 529  
 DB 2529 KEKS--STAMEMLQTLKELNERVALHNDQEAQKAEQNLSQVECLEEKAQLQLD 2586  
 QY 530 AFHINHIYAVLPESV-----QEIGRSAPFRQGANLIFPMGSKVKTGLGEMAFPLSNRLEHD 585  
 DB 2587 EAKNNY---IVLQSSVKGLIQEV-----EDGKQKL-----EKKDBEISRLKNQIQ--- 2628  
 QY 586 LSEQKQLTEIPVOAFSDNAL-KEVLLPASLKTIRREAFKKNHLKQLEVASALSIAFNAL 644  
 DB 2629 --DQEQVLVSKLSQVEGEHQLWKEQNL--ELRNLTVELEQKITQVLQSKNAS-----L 2675  
 QY 645 DDNDGDEQDNKVVVTHNSYALADGEHFTVPDPKLSSTIVDLEKILKLTIEGLDYSTLR 704  
 DB 2676 QD-----TLEVLQSSYKNLENELELTQMDKMSF-----VEKNKM-----TAK 2713  
 QY 705 QT-TQTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAERKALVTKK 763  
 DB 2714 ETELQREMHEN--AQKTAELQDEL-SGEKNRLAGELQLLLEEI-----K 2754  
 QY 764 ATKNGCOLLERSINKAVLAYNNSAIKKA-----NVKRELEKELDLLTGLVEKGFLAQATMVQ 819  
 DB 2755 SSKD-QLKE-----LTENSELKKSLDCHMDHQVEKGVREBEIAYQLRHEAEKKH 2806  
 QY 820 GVYLLKTPPLPEYVYIGLVVYFDK-----SGKLIYALDMSDTIG 858  
 DB 2807 QALLLDTN---KQYEVEIQTYREKLTKEECCLSSQKLEIDLLKSKSEELNLSLKATQIL 2863  
 QY 859 EGQKDA-----YGNPILNVDDNEGHYALAVATLADYEGDLIDIKTLNSKLSQL----- 906  
 DB 2864 BELKKTMDNLKYNQLKENERAQGMKLLIKSKQLE--EKEILQKELSQLQAQEK 2921  
 QY 907 -----TSIRQVPTAAYHRAGIFQAIQNAAAAEQQLPKPG-----THSEKSSSESAN 954  
 DB 2922 QKTGTVMDTKVDELTTEIKE---LKETLEKTKADEYLDKYCSLLISHEKLEKAKEMLE 2978  
 QY 955 SKDRGLQSNPKTNRGRHSAIL 975  
 DB 2979 TQVAHLCSQSKQSDSRGSPLL 2999  
 RESULT 81  
 ADO84842  
 ID ADO84842 standard; protein; 1010 AA.  
 XX ADO84842;  
 AC ADO84842;  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE E faecalis surface anchored LPXTG protein SeqID3.  
 KW LPXTG; cell wall-anchored surface protein; Gram positive bacterium;  
 KW extracellular matrix molecule; sequence database; C-terminal;  
 KW immunoglobulin-like fold region; Ig-like fold region; antibacterial;  
 KW vaccine; gene therapy; infection; medical device; prosthesis;  
 KW premature newborn; AIDS; debilitated cancer; bone marrow transplantation.

XX OS Enterococcus faecalis.  
 XX PN WO2004025416-A2.  
 XX PD 25-MAR-2004.  
 XX PF 15-SEP-2003; 2003WO-US028789.  
 XX PR 13-SEP-2002; 2002US-0410303P.  
 XX PA (TEXA ) UNIV TEXAS A & M SYSTEM.  
 XX PA (INH-) INHIBITEX INC.  
 XX PA (UAB-) UAB RES FOUND.  
 XX PI Hook M, Xu Y, Sillanpaa JV, Sthanam N, Ponnuraj K, Patti JM;  
 XX PI Hutchins JT, Hall A;  
 XX PR WPI; 2004-315694/29.  
 XX PT Identifying LPXTG-containing cell wall-anchored surface proteins from  
 XX PT Gram positive bacteria, for treating infection caused by the bacteria,  
 XX PT comprises searching sequence information database for the sequence having  
 XX PT LPXTG-motif.  
 XX PS Claim 16; SEQ ID NO 3; 96pp; English.  
 XX CC This invention relates to a novel method of identifying LPXTG-containing  
 CC cell wall-anchored surface proteins from Gram positive bacteria that bind  
 CC to an extracellular matrix molecule which comprises searching a database  
 CC of sequence information for a putative protein sequence having the LPXTG-  
 CC motif in its C-terminal region and analysing the sequence for the  
 CC presence of one or more immunoglobulin (Ig)-like fold regions. The  
 CC invention may be useful for the production of compounds with an  
 CC antibacterial activity or for production of a vaccine. In addition the  
 CC disclosed sequences may be useful for gene therapy. The antibody is  
 CC useful for treating or preventing an infection of Gram-positive bacteria  
 CC in a human or animal patient. The method and the proteins are useful in  
 CC generating antibodies for treating and preventing the spread of  
 CC infections of Gram positive bacteria, for interfering with, or inhibiting  
 CC binding interactions by Gram positive bacteria, for monitoring the level  
 CC of gram positive bacterial antigens, or antibodies recognising the  
 CC antigens in a human or animal patients suspected of containing the  
 CC devices and prostheses caused by such organisms, and in treating or  
 CC preventing infections in highly susceptible groups such as premature  
 CC newborns, AIDS and debilitated cancer patients, and bone marrow  
 CC transplantation. The present sequence is that of a surface anchored LPXTG  
 CC protein identified using the method of the invention.  
 XX SQ Sequence 1010 AA;  
 Query Match 3.8%; Score 191; DB 8; Length 1010;  
 Best Local Similarity 19.6%; Pred. No. 0.0087;  
 Matches 215; Conservative 159; Mismatches 391; Indels 334; Gaps 50;  
 QY 111 SGEQLQESLSLNKTPSTSNWEIC--DFITK-----GNTLVGLSKSGVEKLSQTDHL 163  
 DB 30 SGEKLNQSAISLAKQSVAKDQVITVKGFINKGTVGNGNTV-----DDQL 77  
 QY 164 VLPSQAA-----DGTQLIQVASPAFTPKKTAIAEYT 195  
 DB 78 TIPANVAINEKTPSSLTQWDQVTEATSYEVERDGTVPGNITQNTATFDGSPFISEHT 137  
 QY 196 SR---AGENG--EISQ-----LDVDGKEIINEGEVFNYS-----LLKKVT---IPTGY 235  
 DB 138 FRVRAVGKNGVSEWSEPIKGTQDDPYKETINQVATSNLPQPGAEKLLKTDKDLSTGW 197  
 QY 236 KHIGQDAFVDNKNIAEVLNLPESLETISDYAFALALK-----QID-----LPDNLKATG 284  
 DB 198 -----HTNWSTGANPSDGNF--LSLKFGLGAEYQMDKTEYLPFRNAGN 240  
 QY 285 EL-----AFFDNQITGKLSLPQLMRLAERAFKSNHIKTFEPRGN----- 324

DB 241 NILQYRTSKOGANWTFSEPINWKQDALTKTETKQAVRFVEMKVLKSVGNFGSGRE 300  
 QY 325 -----SLKVTIGEASFQDNLSQLMLPGLEKIESEAFTEG--NPGD----- 362  
 DB 301 MLFYKQPGTEGILHGDITNDGTIDENDAMSRYNYTGLESDVSD--FNGYVVERKGDLNKNGVI 359  
 QY 363 DHYNNRVVLWTKSG----- 380  
 DB 360 DAYDISYVLRQLDGGIEIPDVEEIAAGLSLAVVNGKDYLPGLDTLFIKLGQDLKNIN 419  
 QY 381 GLATENTYVNDKSLWQESPEIDYTKWLEEDFTYQKNS-----VTGFSNKGLOKVKRNK 434  
 DB 420 ALSTKMSFDSSKFLVGQPAITNTQOMENYSKYRKHSDNVENLYLVLSNOG-----NK 473  
 QY 435 NLEIPKQNGVTIITEIGNAFENVDFQNKTLRYKYLEEVKLPSTIRKIGAF----- 485  
 DB 474 QL-----LNG-----SMDLVTFKVKVKTETTRVGRATTVEQPLQFDMSSQLLVG 516  
 QY 486 -AFQNNLKSFEASDDLEEIKEGAFMNNRIETLELKD-----LVTIGDAAPHINH----- 535  
 DB 517 QGFQATLSDFSVT-----VKPTELVDKELLQALITLQARVEKEYTPT 561  
 QY 536 --IYAVLPESVQEIERSAFRON-----GANNLI FMGSKVKTLGEMAFLSNRLEHLDLSQ 589  
 DB 562 WAIFKPILDEAVAVLANEQATQTDVSAEAENLEKASOLEKMPDVA---NKAD-LEKAIQ 617  
 QY 590 KOLTEIPV--QAFSDNALK--EVLIPASIKTIRBEAFKKNHLKOLEVASALSIAFNALD 645  
 DB 618 EGLAKKSGDGEFTETTKVJUEESLAAQAQVFAQKVTQEEIDQ----- 661  
 QY 646 DNDGDEQDNKVVVTHNSVALADGEHFIVDPDKLSSTIVDLEKILKLEGLDYSTLRQ 705  
 DB 662 -----ATKTLREAIAQLKEQPVAVDKEITLKEQIAQ--ARGRKEPEGYQFT---K 705  
 QY 706 TTQTFQDMWTAGKALLSKSNLRQGEKOKFLOEQAFFLGRVDLDKAIKAEKALVTKKA- 764  
 DB 706 ETEKQLQEQAIQAAEAIVAKETATKEVSEALNALETAMAQLKEVPLVKNKQDLQEVVVKRAQ 765  
 QY 765 ----TKNQLLERS----INKAVLAYNNSAIKK--ANVKRLEKELDLTLCLVEGKGLQAQ 815  
 DB 766 QVTPSEGHQFTASSLQELQKALLAAXNT--LKNPAANQKXMDIDVAELTSAIDG---LQEE 821  
 QY 816 TMVQGVYLLKPLPLPEYV---IGLVNFDKSGKLIYALDMSDTIGEGOKOAYGNPILNV 872  
 DB 822 VLVTDKKALEAMIAKAKAIKPSAGKEFTSESKARLTAIDQAEGL-LADKNARQEQI--DI 879  
 QY 873 DEDNEGYHALAVATLADYEGLDIKTILNSKLSQLTSLRQVPTAAVHRAG--IFQAIQNA 930  
 DB 880 AEKN-----VKTALD--SLE-EQVLQTDKTKLQKELLQKAEITLKPAGKQFTKASQ 929  
 QY 931 AE---AEQLLPKPGTHSEKSSSESANSK-----DRGLOSNNKTNRGRHSAILPRTG- 979  
 DB 930 ABAIKQAKALVEDPNATQEAVDKCLSIILSQAEIAMEAEPISSNNGNHNHSTVSGTGV 989  
 QY 980 ---SKGSFVYGLGYTSVA 995  
 DB 990 TSQKGATGTTTGTTS 1008  
 RESULT 82  
 ABB60349  
 ID ABB60349 standard; protein; 1216 AA.  
 XX  
 AC ABB60349;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 7839.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmaceutical.  
 XX





CC is useful in determining a loss of expression of any one of the DISC1  
CC partner. The polynucleotide and polypeptides are useful in the  
CC manufacture of a medicament for the treatment of schizophrenia, other  
CC psychotic disorders and mood disorders. The present sequence is used in  
CC the exemplification of the present invention.  
XX  
SQ Sequence 3616 AA;

RESULT 84

ABU43311

ID ABU43311 standard; protein; 3692 AA.

XX

AC ABU43311;

XX

DT 19-JUN-2003 (first entry)

XXII

DE Protein encoded

[illegible]

KW Antisense; prokaryotic essential

**SECRET**

Os staphylococcus epidermidis.

XX  
NW  
WC000277193-22

FN  
WU2002/183-AZ.  
YY

03-0CT-2002

XX  
FD-302-100

PF 21-MAR-2002: 2002WO-US009107.

**XX**

PR 21-MAR-2001; 2001US-

PR 06-SEP-2001; 2001

PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362695P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlson KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
DR  
DR N-PSDB; ACA47181.  
XX  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 3692 AA;

Query Match 3.8%; Score 190.5; DB 6; Length 3692;  
Best Local Similarity 19.9%; Pred. No. 0.057;  
Matches 230; Conservative 184; Mismatches 399; Indels 343; Gaps 56;

QY 30 EPIKQQTQASSISADYABSSGSKLKNINETSFGPDDTVTLDPDKQKTTPEKI--KQNL 87  
DB 2696 ENILPATTVKOKAKADVNAEKEK-LQINS-----NDEATTEELVASDNL 2741  
QY 88 AKGPREQLKAYTE-----NTESEKQI-----TSGSOLEOS--KESLSL 124  
DB 2742 -NHVVETTNQALIEDAPDNNVNEKNGIGTTRDLOPLVVKPTAKSIESAVEKKTEI 2800  
QY 125 NKTVPSTNWEICDFTTKGNTLVGSKGVEKLSQDHLVLPQAAQDGLIQVAFPT 184  
DB 2801 NQTONATHD-EVREGNLQNLQHEKAKDNVQ-SQTNQVQV---ENAEQNSLDQINN--FR 2853  
QY 185 PD---KKTAAIAYTSRAGEN--GEISQ-----LDVQKKEIIN-----E 217  
DB 2854 PDFSKKRNABIV-KAQQNKIDIEIEFBSATQEBEKONALQHLQVKEIINSINQAND 2912  
QY 218 GFVFNFS-----YLLKKVTIPTGKHHGQDAFVDNKNIAEVN-LPESELETIS 262  
DB 2913 NEVDNAKTSGLNITEYREYNKKONALIKLYD-----VSDTQEAINGVPDATE--D 2963  
QY 263 DYAFHLALKQIDLPNLKAIQELAFPDNQITG-----KLSLPR-----QJLM 304  
DB 2964 ELQEANSKLNKI-LLDAKKQIG-LAHTNNEVDVINEYSQKMTILPRVDTKAVARSVLN 3021  
QY 305 RIAERAFKS-----NHK-----TIEFRGSLKVIQGEASFOQNDLS 340  
DB 3022 ALAKQIKITFENTADVTHEERNDANHVKEQLSLVFNAIEKDKRKIQVADQELFGLNELN 3081  
QY 341 QLMPLDLEKIESEFTPGDDHNNRVLVTKSGKNPGLATE---NTYVNDPKSLWQ 397  
DB 3082 SIFINITQKPTARKAISGNA--SQLNNSI-----NNTPYATEERQIALNKVKAIVD 3131  
QY 398 ESPPE-IDYTKWLEEDFTYQKNSVT-----GF-----SNKGQKQKRNKNLEI 438  
DB 3132 DANKEIREANTDSEVLGTGKSNATILLOAISADVQVQKPAFEINAQAEIQERINGSDA 3191  
QY 439 PKHNGVITEIG---DNAFRNVDFQNTLRKYDLEB--VKLPSTIRKIGAFQFOSNLK 493  
DB 3192 TREKKEALKQVDTLVNHSFIFINNKNQVEYDQDKTIEAHKIPKISTIKPOLALNEI 3251  
QY 494 SPEASDDEEIKEGAFMNNRIETLEK---DKLVITIG-----DAAPHNHIAIVLP 542  
DB 3252 TIQLDTQDRLIK-----NKESTVEKASAIKLIKTAARIAESIDKA-QTNEEVNKK 3305  
QY 543 ESQVQIGR---SAFRQNGANNLIFMGSKV-----KTIGEMAFLSNRLB----- 582  
DB 3306 QSIDBISKILPVIEIKSARNEIHOQAEVIRGLINDNEATKEEKDIALNQLDITLTQAN 3365  
QY 583 -HLD-----LSEQKQLTEIPVQAFSDNALKEVLLPASLKTIREAPFK-NHLQLEVA 633  
DB 3366 VSIDQALNNEANRAKBTANSEINKISVIAIKK---PEAIAEIQBLADKCLKKPKQSEA 3422  
QY 634 S-ALSHIAPNALDDGDGDFQFNKVVVTKHNSYALADGEHIVDPDKLSSTIVLEKIL 692  
DB 3423 TIEEQSALNELEQ-----ALKSAINHIHOSQNNESVSAALKESI 3462  
QY 693 KUEGLDYSTLRQTQTQTFDMMTAGKALLSKNLRQGEKQKFLQEAQFFLGRVLDKAI 752  
DB 3463 SLIDSIEIQAHKLEAKAYIDGYSDDKINDISSRATNEEKQIFVSKLALINRT--HKQI 3520

QY 753 AKAEKALVTYKATKNGQLLERSINKAVLAYNNSAIKKANV-KRLEKELDLTLGLVEGKGP 811  
DB 3521 DEATFTVSVETIVRFKVEADKLSIVR-----KKAKASKEIELEADHVQKQMINAN-- 3571  
QY 812 LAQATMVQGVVLLKTLPLPLPEYIYGLNVYFDKSGKLIYALDMSDTIGEGQKDAYGNPILN 871  
DB 3572 LSASTRVK-----QNAITLINEIVSNALSQ-----LN 3598  
QY 872 VDENEGHALAVLATLADYEGLDIKTILNSKLSQTSITRQVPTAAHYHRAGIFQAIQNA 931  
DB 3599 KVTNNK-----EVDEIVNETIEKLSIQ-----IR 3623  
QY 932 EAEOLLPPGTHSEKSSSESANSKDRGLOSQNPKNTRGRHSAILPRTGSKGSFVYIGILY 991  
DB 3624 EDKILSQRSTSTWTEKSNQCYSSENNTIKSLPEA--GNADKSLP-----LAGV 3670  
QY 992 TSVALLSLITAIKKKK 1007  
DB 3671 TLISGLAIMSRKKKK 3686

#### RESULT 85

ADN22762  
ID ADN22762 standard; protein; 925 AA.

AC ADN22762;

XX 02-DEC-2004 (first entry)

DE Bacterial polypeptide #5415.

XX Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KW bacterial polypeptide.

OS Bacteria.

XX US2003233675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

XX (HINK/) HINKLE G J.

XX (SLAT/) SLATER S C.

XX (CHEN/) CHEN X.

XX (GOLD/) GOLDMAN B S.

PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
PT microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 5415; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a  
CC promoter functional in a plant cell, where the promoter is positioned to  
CC provide for expression of a polynucleotide encoding a polypeptide from a  
CC microbial source. The invention also relates to a transformed plant  
CC comprising the recombinant DNA construct and a method of producing a  
CC transformed plant having an improved property. The plant is a crop plant  
CC such as maize or soybean. The method of producing a transformed plant  
CC having an improved property comprises transforming a plant with the





the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 2398 AA;

Query Match 3.7%; Score 189; DB 6; Length 2398;  
Best Local Similarity 18.7%; Pred. No. 0.039;  
Matches 210; Conservative 153; Mismatches 444; Indels 314; Gaps 46;

QY 2 KKHKTVALTLTVSVVTHN-----QEVFSLVKEPIKQTOASSISGDAVSSGSKG 55  
DB 1111 KQHLG-----TLTHITTAQRNDLTNQISOATNLGAVESVQKN--ANSLDGA-----MGNLQ 1159  
QY 56 LKINETSGBV-----DDTVTDLFSDKRTTPEKID-----NLAKGPREQELKAV--- 99  
DB 1160 TAINDKSGTLASQNFILDADEQKRNAYNQAVSAETILNKQTGTPNTAKTAVEQALNNVNA 1219  
QY 100 -----TENTESEKO-----ITSGSOLEQ-SKESL---SLNKTVPTSINWEICDFTKGT 145  
DB 1220 KHALNGTQNLNNAKQAATAINGASDLNQKQDALKAKAQAQAVSNAQDVQHNATLNT 1279  
QY 146 LVGLSKSGV-----EKLSDTHLV-----LPSQAADGT 173  
DB 1280 AMGLKHAADKNTLASSKYVNAUSTKQNAVTKVNAEHLISGTPVTVPSEVTAA 1339  
QY 174 QIUQVAFAPTPD-----KKTAAIAYTSRAGENGESISOLDVD 210  
DB 1340 NOVNSAKQELNGDERLREAKQNANTAIDALTQNTTPQAKLKEQVQANRLEDVQTVQTN 1399  
QY 211 GKELINEGEVFNLSYLLKKVTIPTGYKHIGQDAFVDNKN-----IAEVNLPESL 258  
DB 1400 GQALNNAMKGLRDSIANETTVKTSQNY--TDASPNQSTYNSAVSNKAGIINQTNPN--- 1454  
QY 259 ETISDYAFHAHLKQIDLPDLNKAIGELAFQNFQITGKLSL-----PRLMLRAERAF 311  
DB 1455 -TWDTSAITQATOVNNAKNGLAENLR--NAQNTAKQNLNTLSHLTNQKSAISSQID 1511  
QY 312 KSNHKTIFBFRGNLSKVIGESAFQNDLSQLMPDGLKEISEAFT--GNPGDDHYNNRV 369  
DB 1512 RAGHVSEVATKNAAT---ELNTQMGNLQEAHQNTVK-QSVKFTDADKAKRDATYNAV 1567  
QY 370 ----VLWTKSGKNPSGLATENTYVNPDKLSLWQESPEIDYTKWLBEDFTYQKNSVTGSN 424  
DB 1568 SRAEAILNKTOGANTSQDVAAAIQNVSSA-----KXALNGDQ 1606  
QY 425 KGLQKVRKNLEIPQKHNGVTITEIGDNAFRNVDPQNKTLRKYDLEEVKLPSTIRKIGA 484  
DB 1607 -----VTNAKN-----AAKVALNNTSINNAQRDLTKIDQATVAGVBA 1647  
QY 485 FAFQSNLKSFRASDDLEEBKEGAFNNRIETLEK-----DKLVTIGDAFHHNY 537  
DB 1648 VENTSTQLNTAMAN-----LQNG--INDKNTLASENYHDADSKTAYTQAVTNAENIL 1700

QY 538 -----AIVLPESVOEIGRSAFRQNCANNLIIFWGSVKVTKLGEMAFLSNLEHLDLSE 588  
DB 1701 NKNSGSLDKTAVENLQVANAQKALNGNHLNLEQAKGNANT-----TINGLQHLTTTAA 1754  
QY 589 QKQLTEIPVQAFSDNALKEVLLPASL-----KTIREEAFKKNHLKQLEVASALS 639  
DB 1755 KDKLQKQVQQAQNVAGVDTVKSSANTLNGAMGTCLNSIQDNTATKNGQNYLDATERNKTN 1814  
QY 640 AFNALDDNDG-----DEQFNKVVVVKTHNSYALADGEHFIVDPDKLSSTIVD-- 687  
DB 1815 YNNVDSANGVINATSNPNMDANINQIATQVTTSTKNAL-DGTHNLQAKQATATNAIDGA 1873  
QY 688 --LEKILK-----LIEGLD-----YSTL 703  
DB 1874 TNLNKAQKDALKAQVTSQORVANVTISIQQTANELMTAMQLOHQHGIDDENATQTKYRDA 1933  
QY 704 RQTTQTPFRDMTTAGKALLSK---SNLRQGEKQKFLQEA---QFFLGRVLDLDAKAAE 756  
DB 1934 EQSKKTAYDQAVAAAKAILNKQTSNSDKAAVDRLAQVTVSTKDALNGDKAKLAEAKAAK 1993  
QY 757 KALVTKKATKNGQL--LERSINKAVLAYNNSAIKANVKRLEKELDLTLGLVEGKGLAQ 814  
DB 1994 QNLGTLNHTNAQRTDLEQINQATTVDGVNTV-KTNANTLIDGAMNLSQGSINDK----D 2048  
QY 815 ATMVQGVVLLKTPPLPEYIYGLNVVFDKSGKLIYALDMSDTIGEGQXDAGNPIILNVDS 874  
DB 2049 ATRLNQNYL-----DADESKRNATQVATAEGILNKQTGGN-TSKADV 2091  
QY 875 DNEGHALAVATLAD--YEGLDIKTILNSKLSQTSIRQVP 913  
DB 2092 DN-----ALNAVTRAKAALNGAD--NLNRAKTSATNTIDGLP 2126  
RESULT 88  
ABP29526  
ID ABP29526 standard; protein; 1240 AA.  
XX AC ABP29526;  
XX DT 02-JUL-2002 (first entry)  
XX Streptococcus polypeptide SEQ ID NO 8228.  
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX group A streptococcus; Streptococcus pyogenes; antibacterial;  
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX Streptococcus pyogenes.  
XX WO200234771-A2.  
XX PD 02-MAY-2002.  
XX 29-OCT-2001; 2001WO-GB004789.  
XX 27-OCT-2000; 2000GB-00026333.  
XX 24-NOV-2000; 2000GB-00028727.  
XX 07-MAR-2001; 2001GB-00005640.  
XX (CHIR-) CHIRON SPA.  
XX (GENO-) INST GENOMIC RES.  
XX Telford J, Masignani V, Margarit Y Rosl, Grandi G, Fraser C;  
PI Tettelin H;  
XX WPI; 2002-352536/38.  
XX DR N-PSDB; AEN70157.  
XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
XX





PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voess H;  
XX  
DR WPI; 2002-010914/01.  
XX  
XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
PT and prevention of *Listeria* and related bacterial infections, and related  
PT polypeptides.  
XX  
XX Claim 6; SEQ ID NO 456; 192pp; French.  
XX  
XX The present invention relates to the genome sequence of *Listeria*  
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in *L.*  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of *L. monocytogenes* and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate *L. monocytogenes*-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccines compositions for the treatment or prevention of infections by *L.*  
CC monocytogenes and related organisms. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 2013 AA;

Query Match 3.7%; Score 188.5; DB 5; Length 2013;  
Best Local Similarity 21.9%; Pred. No. 0.032;  
Matches 250; Conservative 148; Mismatches 429; Indels 315; Gaps 66;

1385 LNVAAATDTHDGNITPVWDYSKVKDVLGTPVTVTATDASGNKATQTVNLRIVDTTSPTI 1444  
QY 465 -----LRKYDLEEVKLPSTIRKIGA-----FAFQSNLKS 494  
DB 1445 LITNPLTVSTENMERK--LTBOELYTAAGLIGDNDYDLAPQAVQPNQPMWFTSNFSTI 1502  
QY 495 FEASDDLEIEKEGAFM-----NNRIE---TELEKDKLVITIGDAAFHINHIYAIYL 541  
DB 1503 F---GDIASVRAGOVQVNLADSSGNAIPQTIINVD---TVGPVKADNVSHVNT 1556  
QY 542 PESVQIEGRSAPRONGANNLIPMGSKVTKLGMFAFLSNRLEHLJLSE--OKOLTEIPVOA 599  
DB 1557 TKTEAEFFQDARLDVTDNN-----DVTTLTIITSNFAEKVNLNPKGVEVITATDT 1608  
QY 600 PSDNALKEVILLPASL-KTIREAPKKNHLKQLEVASA--LSHIAFNALDDNDGDQFQDNK 656  
DB 1609 KGNQTKSITVQVSKDKPVTADPKISYOGKIEVTEANFLSGVHAETDELGDVKITSD 1668  
QY 657 VVVVTKTHN---SYALA---DGEHFIVDPDKLSSITVDLEKILKILIEGLDYSTLRQTQT 709  
DB 1669 FAEKVDFNKGVTYTVTLNAKDEYGNTPAEVKVSVSI--FNKIAPTFNADNKTIEAVNEL 1726  
QY 710 QFRD--MTTAGKALLSKNLQGEKQKFLQEAQFFLGRVLDLDAIAKAEKALVTKATK- 766  
DB 1727 PSLESIFKIEAKDYLSGNKL---KVTYTPK--QTIKGNVPGEYSI-----KVTTKD 1772  
QY 767 -NGOLLERSINKAVLAYNNSAIKKA-NVKRLE---KELDLL-----TGLVEG---KGP 811  
DB 1773 DSGNIAETVTLTIKDTGTPSKMTKSKLEQVSKPEFNWITFFGIKATDIVDGDVTKNI 1832  
QY 812 LAQATMVQ---GVYLLKTPLPPEYIYICLVNVPYFDKSGKLIYALDMSDTIGEGQDAYGN 867  
DB 1833 KVDSEVNLNKGVTY-----PIYFTVTDALGNESKLSKSTVQIVDT-----SS 1875  
QY 868 PILNVEDNEGYHALAVATLADYEG-LDIKT-ILNS--KLSQTSIRQVPTAAHYHRAGIF 923  
DB 1876 PELTIDKLEISYP--TGKTISDKQLQDIGTKVTSYGVTKVTTNLSKI--VDMNKAGKY 1931  
QY 924 QAIQNA-----AAAEQOLLKPGGTHSEKSSSESANSKDRGLQSNPKTNRGHSAIPLR 977  
DB 1932 KVTVTATNSSGGVAEKITILL-----TVKNSDSFSIAVPSKD-----DNKVKPAKNIPK 1979  
QY 978 TG 979  
DB 1980 TG 1981

RESULT 90  
ABU32684  
ID ABU32684 standard; protein; 2013 AA.  
XX  
XX AC ABU32684;  
XX  
XX DT 19-JUN-2003 (first entry)  
XX  
XX DE Protein encoded by Prokaryotic essential gene #18211.  
XX  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
XX OS *Listeria monocytogenes*.  
XX  
XX PN WO200277183-A2.  
XX  
XX PD 03-OCT-2002.  
XX  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX  
XX PR 21-MAR-2001; 2001US-00815242.  
XX  
XX PR 06-SEP-2001; 2001US-00948993.  
XX  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX  
XX PR 08-FEB-2002; 2002US-00072851.  
XX  
XX PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykend JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA36554.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 60608; 1766pp; English.  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from Wipo at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2013 AA;  
 Query Match 3.7%; Score 188.5; DB 6; Length 2013;  
 Best Local Similarity 21.9%; Pred. No. 0.032;  
 Matches 250; Conservative 148; Mismatches 429; Indels 315; Gaps 66;  
 QY 18 VTHNOEVLVKEPIKOTQASSISGADYABSSGSKLKNETSGPVDDT--VTDL--- 72  
 DB 975 IYLFNAIVPTDAPNDEIAYNSIAYRIDKGTSGTSLKASEPPRGVKSNTSPDNLIA 1034  
 QY 73 ----\$GDKRTTPKIKDNLAGPREQLKAV-----TENTESEK---QITSGSQLEQSK 120  
 DB 1035 GNSFSD-----LNKNGVKDADELGLNAVKLDLYKONNEPEKVTYVTTSSDALDK 1087  
 QY 121 SLSLN-----KTVPSISNWEICDPIKNG--NTLVGLSKG-----VEKLSQTDHLVLP 168  
 DB 1088 LFDNGLNNGTVTKIAAHLN--KNADFITTPGNKIVKDKSDSIGWITVNNSTFTI--DD 1144  
 QY 169 AADGTQL-----IQVASFAPTP-----DKTAIAEYTSRAGENGESQLDDVGKEI 214  
 DB 1145 LANGNPKNLVGIQIPIYMATPIKGVAVFNKDGEPKITSY-----QDYEIFALFDKQKEV 1200  
 QY 215 -----INEGV--FNSYLLK-----KVTIPTGYKHIGQDAFVDNKNIAEVLNPLESLET 260  
 DB 1201 QSAIKNTSKGPFSDVNAIKPADFKLVKTAFTGTNFV-----YSKKN----- 1243

QY 261 ISDYAPAHALAKQIDLPNLKAIGELAFFNQITGKLSLPQLMRLAERAFKSNHIKIE 320  
 DB 1244 ----PLFNMSTKTYTLANSVPGVGVA--EIIYIT-ETSKPTTKI--ILDKAVTPNAL-TIE 1294  
 QY 321 PRGNLSKIVIGASPDNDLSQLMLPDGLEKTESAFTGN-----PGDDHYNNRVVLTWTKSG 376  
 DB 1295 SSDEATEVTNWTLESDSGTVVY-----TGIGNTIRIPNDE-----GTVIA 1335  
 QY 377 KNPGLATENTYVNPDKSLMOESPEIDY---TKWLEBEDFTYKNSVTFSGSKNGKQKVRN 433  
 DB 1336 KNTATDEAGNT-ASDEKTF-----DIDYVPTLVNQDASAEVNSTE--ANIGWIK---P 1384  
 QY 434 KNLSEIPKQNG-----VTITEIGNA-----FNVDPQNKNT- 464  
 DB 1385 LNVAATDTHGNIITPVVDYSKVKWDVLGTPVTVTATDASGNKATQTVNLRIVDTTSPTI 1444  
 QY 465 -----LRKYDLEBVKLPSTIRKIGA-----PAPOSNNLKS 494  
 DB 1445 LITNPLTYSTENWRK--LTSQELYTAAGLIGDNDYDLAPQAVQPNKQPMVFTSNFSTI 1502  
 QY 495 PEASDDLEIEKEGAPM-----NNRIE---TLEKDKLVITIGDAAFHNIHIAVL 541  
 DB 1503 F---GDIASVRAGQYQVQVNLADSSGNOAIPQTITINVVD---TVGPVIKADNVSYHNT 1556  
 QY 542 PESVQIEGRSAFRQNGANNLIFMGSKVKTGEMAFNLNRLEHLDSE--OKQLTEIPVQA 599  
 DB 1557 TKTEARFPQDARLOVDTNN-----DDTVDLIITSNFAEKVNLNPKGYEVITATDT 1608  
 QY 600 FSDNALKEVLLPASL--KTIREAFKQNLKOLEVASA--LSHIAFNALDDNDGDEQFONK 656  
 DB 1609 KGNQTTKEITVOVKDKPVITADPKISVQKIEVTEANFLSGVHAETVDELGDGVKITSD 1668  
 QY 657 VVVKTHN---SYALA---DGEHPIVDPDKLSSTIVDEKILKLEGLDYSTLRQTQT 709  
 DB 1669 FAEKVDENKVGTYTTLNAKDEYGTAEAPVKVSVSI--FNKIAPTFFNADNKTIEAVNEL 1726  
 QY 710 QFRD--MTTACKALLSKNSLFGQSKQKFLQEAQFPLGRVDLDKAIKAEKALVTKKATK- 766  
 DB 1727 PLSLSIFKIEAKOYLSGNKL-----KVYTPPE-QTIKGNVPGEYSI-----KVTTKD 1772  
 QY 767 -NGQLLERSINKAVLAYNNSAIKKA-NVKRLE---KELDILL-----TGLVEG---KGP 811  
 DB 1773 DSGNIAETVTLTIKDTTGPSIKMTKSTKLEQSKEPNWTTPFGIKATDIVDGDVTKNI 1832  
 QY 812 LAQATWQ-----GYVLLKTPPLPEYIGLVNVPYKSKLIYALDMSDTIGEGQKDAYGN 867  
 DB 1833 KVDSEVNLNKVGTY-----PIYFTVTDALGNSSKLSKSTVQIVDT-----SS 1875  
 QY 868 PILNVDEDEGYHALAVATLADYGL-DIKT-ILNS--KLSQLSIRQVPTAAVHRAGIF 923  
 DB 1876 PELTIDKLEISYP--TGKTIQSDKQFLQDIGTKVTSYGVTKVTNNLSKI--VDMNKAGKY 1931  
 QY 924 QAIONA-----AAEAEQLLPKPGTHSEKSSSSSSANSKDRGLQSNPKTNRGHSAILPR 977  
 DB 1932 KVTYATNSSGGVAEKITILL---TVKNSDSSFIATVPSKD-----DNKNKPAKNIPK 1979  
 QY 978 TG 979  
 DB 1980 TG 1981  
 RESULT 91  
 ADA89551  
 ID ADA89551 standard; protein; 1448 AA.  
 XX  
 AC ADA89551;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Staphylococcus aureus antigenic protein #90.  
 XX  
 KW antigenic protein; vaccine; Staphylococcus aureus; pathogenic organism;  
 KW antibacterial; neuroprotective; immunosuppressive; antiinflammatory;

KW antiulcer; immunostimulant; ophthalmological; pathogenic microbe;  
 KW bacteraemia; septic shock; organ infection; skin infection;  
 KW bacterial basal colonisation; bacterial eye infection; septicaemia;  
 KW tuberculosis; food poisoning; blood infection; peritonitis; endocarditis;  
 KW sepsis; meningitis; pneumonia; stomach ulcer; gonorrhoea; toxic shock;  
 KW necrotising fasciitis; impetigo; histoplasmosis; Lyme disease;  
 KW gastro-enteritis; dysentery; shigellosis; skin disorder.  
 XX Staphylococcus aureus.

OS WO2003011899-A2.

PN 13-FEB-2003.

PD 02-AUG-2002; 2002WO-GB003606.

PF 02-AUG-2001; 2001GB-00018825.

PR 09-JAN-2002; 2002GB-00000349.

XX (UYSH-) UNIV SHEFFIELD.

PA (BIOS-) BIOSYNEXUS INC.

XX Foster S, Mond J, Clarke S, McDowell P, Brummel K;

PI WPI; 2003-256434/25.

DR New antigenic polypeptides from Staphylococcus aureus or S. epidermidis,

XX useful as a vaccine for immunizing humans against e.g. bacteremia, septic  
 PT shock, septicaemia, tuberculosis, meningitis, pneumonia, gonorrhoea or  
 PT impetigo.

XX Claim 4; Page 140-141; 189pp; English.

XX The present invention describes an antigenic protein or its part, which  
 CC is for use as a vaccine. The antigenic protein is encoded by an isolated  
 CC DNA molecule of Staphylococcus aureus or S. epidermidis partial gene  
 CC sequences (designated dnaSA and dna SE, respectively; and which encodes a  
 CC protein expressed by a pathogenic organism. Also described: (1) a vaccine  
 CC composition comprising at least one antigenic protein; (2) a method of  
 CC immunising an animal against a disease or condition caused by a  
 CC pathogenic microbe by administering the antigenic protein or the vaccine;  
 CC (3) an antibody or its binding part obtainable by the method above; (4)  
 CC preparing a hybridoma cell line producing monoclonal antibodies; (5) a  
 CC hybridoma cell line produced by the method of (4); and (6) identifying  
 CC opsonic antigens expressed by a pathogenic microbe. The antigenic  
 CC proteins have antibacterial, neuroprotective, immunosuppressive,  
 CC antiinflammatory, antiulcer, immunostimulant and ophthalmological  
 CC activities, and can be used in vaccines. The antigenic proteins or  
 CC vaccines can be used for immunising an animal (specifically a human)  
 CC against a disease or condition caused by a pathogenic microbe, e.g.  
 CC bacteraemia, septic shock, organ infection, skin infection, bacterial  
 CC basal colonisation, bacterial eye infections, septicaemia, tuberculosis,  
 CC bacteraemia-associated food poisoning, blood infections, peritonitis,  
 CC endocarditis, sepsis, meningitis, pneumonia, stomach ulcers, gonorrhoea,  
 CC strep throat, streptococcal-associated toxic shock, necrotising  
 CC fasciitis, impetigo, histoplasmosis, Lyme disease, gastro-enteritis,  
 CC dysentery, shigellosis, S. aureus-associated septicaemia, food-poisoning,  
 CC skin disorders, S. epidermidis-associated septicaemia, peritonitis or  
 CC endocarditis. The present sequence represents a S. aureus antigenic  
 CC protein sequence from the present invention.

XX Sequence 1448 AA;

Query Match 3.7%; Score 187.5; DB 6; Length 1448;

Best Local Similarity 19.3%; Pred. No. 0.024;

Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;

QY 12 LTTVSVVTHNQEVFSLVKEPILKQTCASSISGA-----DYAESGSKLXINETSGPV 65

DB 140 ITAEETAAANADVNAV-----TQANSNIEANSQNDVDQAKTTGNSI-----183

QY 66 DDTVTDLFSDKRTTPKPKIDNLAQPREQELKAVTENTESEKQITSGSQLEQKESLSLN 125

Db 184 -DQVPTVNNKKATARNEITAIL--NNKLEIQATPDATDEEKQ-----AADAENTENGKAN 237

QY 126 KTVPTSNWEICDFITKGNLTGLSKSGVKEKLSQTDHLVLPQAADGTQLIQVASFATP 185

Db 238 QAISATTT-----NAQVDEAKANAE-----AANAVATP 265

QY 186 D--KKTATAEYTSRAGEEISQLDVGKEIINEGEVFNFSYLLKKVTIPTGYKHGQDAF 243

Db 266 KVVKKQAQKD-----EIDQLQATQTNVNN-----DQDAT 295

QY 244 VDNKNAIEVNLPSLETISDYAFALHALKQIDLPNKLKAIGELAFDNOITKLSLPRQL 303

Db 296 TEEK-----EAAIQQLATAVTDKKNITA-----ATDDNGV-----326

QY 304 MRLAERAFKSNHIKTIEFRGNSLKVIGEASFODNLSQLMLPDGLEKIESEAFTPNGPDD 363

Db 327 -----DOAKGAKNSIQSTQPATAVKSNKNDVDQAVTTQNOAID 366

QY 364 HYNRRVVLWTKSGKPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFS 423

Db 367 -----NTTGATTEK--NAAKDLVLKAKEKAYQDILNAQTT---NDVTQIK 407

QY 424 NKGLOKVRKNKLEIPKQHNGV-----TTEIGDNARFNVDFQNKTL--RKYDLEEVKLS 477

Db 408 DQAVADIQ-----GITADTTIKDVADELATKANEQKALIAQTADATTEKEQ 455

QY 478 TIRKIGAFAPQSN-NLKSFEASDDLEETKEGAFM-----NNRIETL-ELKDK 522

Db 456 ANQQVDAQLTQGNQNIENAAQSIDDVNTAKDNAIQAIIDPIQASTDVKTNRABELLTEMQNK 515

QY 523 LVTIGDAAFPHNHYAIVLPESVQBIG--RSAPFRQNGANNLIFMGSKVKVTLGEMAFLSNR 580

Db 516 ITEILLNNETNE-----EKGNDIGPVRAAY-EGLNNI---NAATTG-----555

QY 581 LEHDLBSQKQLTEIPVQAFSNALKEVLLPASLTIREEAF-KKNHLKQL-----630

Db 556 -----DVTTAKOTAVQVQQLHANPVKK---PAGKELDQAAADKKTQIBQTPNASQOEIN 608

QY 631 -----EVSANLSHIAFNALDDNDGDFQFNKV-----VVKT--HHSNYALADGHP--674

Db 609 DAKQEVDTLNOAKTN--VDQSTNBYVDNAVEKGAKINAVKTFSEYKKDALAKIEDAYN 667

QY 675 ---IVDPD---KLSTIVDL-EKILKIEGLDYSLRQTTQ-----TQFRDMT-715

Db 668 AKVNEADNSNSTSSSIEAEKQKLAELKQADQNVNQATSKDDIEVQIHNDLNDINDYTI 727

QY 716 TAGKA-----LLSKSNLRQGEKQKFL---QEAQFFLGRV-----D 747

Db 728 PTGKKESTTDLAYADQKKNISADTNTAQDEKQQAIKQVDQNVQTALESINNGVDNGD 787

QY 748 LDKATAKAEKALVTKKATNGQLLERSINKAVLAYNNSAIKANVUKLEKELDILTG---804

Db 788 VDDALTQGAAT--DAIQVDAVTVPKANQAI-----EVKAEDTKESIDQSDQJTAEBK 838

QY 805 -----LVEGKGPLAQATMGVGVLLKTPPLPEYIIGL---NVYFKSGK---846

Db 839 TEALAMIKQITDOAKQGITDATTAEVKAQAQ-----GLEAFDNIQIDSTEKQKAI 890

QY 847 ---LIYALDM-----SDTIGEGQKDAYGNPILNVDENEGYPHALAVATLADY-EGLDI 895

Db 891 BELETALDQIEAGVNVNADATTE-EKEAFTNAL-----EDILSKATEDISDQTTNABI 942

QY 896 KTLINLSKLSOLTSIROVETAYVHRAGI-----FOAIONAAAE-----934

Db 943 ATVKNSALEOLKAQRIINPEVKKNALAEIREVVNKQIEIKKNADADASAKEIARTDLGRYF 1002

QY 935 ---QLLPKPGTHSEKS-----SSSESANSKDRGLQSNPKTNRGRHSATL 975

Db 1003 DRFADLKTQTNVAEVLQNTVIPAIEAIVPQNDPDANDTNGDNDNATANSANATP 1062

QY 976 PRTG 979

Db 1063 ENTG 1066







Qy	976	PRTG	979
Db	2093	ENTG	2096
RESULT 94			
ABU19002	ABU19002 standard; protein; 2478 AA.		
AC	ABU19002;		
XX	XX	XX	XX
DT	06-MAR-2003	(first entry)	
XX	XX	XX	XX
DB	Pathogen specific antigen related staphylococcal protein SEQ ID No 220.		
KW	Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;		
KW	hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;		
KW	auto-immunity; vaccine; staphylococcal infection; antibody; cancer;		
KW	autoimmune disease; HIV; hepatitis.		
OS	Staphylococcus sp.		
XX	XX	XX	XX
PN	WO200259148-A2.		
XX	XX	XX	XX
PD	01-AUG-2002.		
XX	XX	XX	XX
PF	21-JAN-2002; 2002WO-EP000546.		
XX	XX	XX	XX
PR	26-JAN-2001; 2001AT-00000130.		
XX	XX	XX	XX
PA	(CIST-) CISTEM BIOTECHNOLOGIES GMBH.		
PI	Mainke A, Nagy E, Von Ahseu U, Klade C, Henics T, Zauner W;		
PI	Minh DB, Vytvytska O, Eitz H, Dryla A, Weichhart T, Hafner M;		
PI	Templemaier B;		
XX	XX	XX	XX
DR	WPI; 2003-075410/07.		
XX	XX	XX	XX
PT	Identifying, isolating and producing hyperimmune serum-reactive antigens		
PT	from a pathogen, for preparing vaccine or medicament for treating or		
PT	preventing e.g. staphylococcal infections, comprises providing antibody		
PT	preparation.		
XX	XX	XX	XX
PS	Example 7; Page 186; 252pp; English.		
XX	XX	XX	XX
CC	The invention relates to a novel method for identifying, isolating and		
CC	producing hyperimmune serum-reactive antigens from a pathogen, tumour,		
CC	allergen, a tissue or host prone to auto-immunity, where the antigens are		
CC	used in a vaccine, comprises providing antibody preparation from a plasma		
CC	pool of a type of animal, or individual sera with antibodies against the		
CC	specific pathogen, tumour, allergen, tissue or host prone to auto-		
CC	immunity. The hyperimmune serum-reactive antigens comprising any of the		
CC	62 sequences of 53-2261 amino acids fully defined in the specification,		
CC	or their hyperimmune fragments are useful for the manufacture of a		
CC	pharmaceutical preparation, particularly a vaccine against staphylococcal		
CC	infections or colonisation against <i>S. aureus</i> or <i>S. epidermidis</i> . The		
CC	preparation of antibodies is useful for the manufacture of a medication		
CC	for treating or preventing staphylococcal infections or colonisation		
CC	against <i>S. aureus</i> or <i>S. epidermidis</i> . The antibody preparations may also		
CC	be used for diagnostic and imaging purposes. Other conditions that can be		
CC	treated include cancer, autoimmune diseases or infections caused by viral		
CC	(e.g. HIV, hepatitis A, B or C), fungal or protozoan pathogens. This		
CC	sequence represents a staphylococcal protein relating to the method for		
CC	identifying and producing pathogen specific antigens of the invention		
XX	XX	XX	XX
SQ	Sequence 2478 AA;		
Query Match 3.7%; Score 187.5; DB 6; Length 2478;			
Best Local Similarity 19.3%; Pred. No. 0.05;			
Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;			
Qy	12	LTTVSVVTHNQEVFSLVKEPILKQTQASSISGA-----DYASSSGSKSLKINETSQPV	65

Db 2033 DRFADKDKTQTNVAVELQNVTTIPAEIAIVPQNDPDANDTNGINDNDATANSNANATP 2092

QY 976 PRTG 979

Db 2093 ENTG 2096

## RESULT 95

ABM71899

ID ABM71899 standard; protein; 2478 AA.

AC ABM71899;

XX 20-NOV-2003 (first entry)

XX Staphylococcus aureus protein #1139.

XX Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;

XX enzymatic assay; antibiotic target.

XX Staphylococcus aureus.

XX WO200294868-A2.

XX 28-NOV-2002.

XX 27-MAR-2002; 2002WO-IB002637.

XX 27-MAR-2001; 2001GB-00007661.

XX (CHIR-) CHIRON SPA.

XX Masignani V, Mora M, Scarselli M;

XX WPI: 2003-120786/11.

XX N-PSDB; ACF73459.

XX New Staphylococcus aureus protein, useful as a vaccine for treating or  
PT preventing Staphylococcal infection, specifically an infection caused by  
PT S. aureus, e.g. sepsis.

XX Claim 1; SEQ ID NO 2278; 49pp; English.

XX The invention relates to novel genes and encoded proteins from  
CC Staphylococcus aureus. A composition comprising the S. aureus protein, a  
CC nucleic acid encoding the protein, or an antibody to the protein, is  
CC useful as a pharmaceutical, particularly as a vaccine for treating or  
CC preventing infection due to Staphylococcus bacteria, specifically an  
CC infection caused by S. aureus. The composition is particularly useful for  
CC treating or preventing sepsis in a patient. The composition can also be  
CC used for diagnostics. The protein is also used in an assay for enzymatic  
CC studies and as a target for antibiotics. This sequence represents one of  
CC the novel S. aureus proteins of the invention

XX Sequence 2478 AA;

Query Match 3.7%; Score 187.5; DB 6; Length 2478;

Best Local Similarity 19.3%; Pred. No. 0.05;

Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;

QY 12 LTVSVVTHQEVFLVKEPILKQTSASSISGA-----DYAESSGSKLKINETSGPV 65

Db 1170 ITAEIIAANADVDAV-----TQANSNIEAANSQNDVDQAKTTGENSI----- 1213

QY 66 DTVTDLDFSKTTEPEKIDNLAQPREQELKAVTENTSEKQITSGSQLEQSKESLSLN 125

Db 1214 -DQVPTVNNKATARNETAIL--NNKLOEIQTDPATDEEQ---AADAENTENGKAN 1267

QY 126 KTVPTSNWEICDFITKGNLVLGSKSGVEKLSQTDHLVLPQADGTQLIQVAFPTP 185

Db 1268 QAISAATT-----NAQVDEAKANAE-----AANAVTP 1295

QY 186 D--KKTAAEYTSRAGENGESQLDVGKKEIINEGEVFNFSYLLKKVTIPTGYKHGQDAF 243

Db 1296 KVVKKQAAKD-----EIDQLQATQTNVNN-----DONAT 1325

QY 244 VDNKNIAEVLNPESLETISDYAFALHALKQIDLPNKAIGELAFDFNOITKGLSIPROL 303

Db 1326 TEEK-----EAAIQQLATAVTDAKNNITA-----ATDNGV----- 1356

QY 304 MELAERAFKSNHIKTIIEFRGNSLKVIGBASFOQNDLSQLMLPDGLEKIESEAFNGPDOD 363

Db 1357 -----DQAKDAGKNSIQSTQPATAVKSNAKNDVDQAVTTQQAID 1396

QY 364 HYNRRVVLWTKSGKNPSGLATNTVNPDKSLWQESPEIDYTKYLEEDTYQKNSVTGFS 423

Db 1397 -----NTTGATTEEK--NAAKDILVLAKEKAYQDILNAQT--NDVTQIK 1437

QY 424 NKGLOKVKRNKLEIPKQHGV-----TITEIGDNAPRVNDFQNKTL--RKYDEEVKLP 477

Db 1438 DQAVADIO-----GITADTTIKDVADELATKANEQKALIAQTADATTEKEQ 1485

QY 478 TIRKIGAFAPQSN-NLKSFEASDDLEIEKEGAFM-----NNRIETL-ELKDK 522

Db 1486 ANQVDAQLTQGNQNIENASQIDDVNTAKNAIQAIQIDPIQASTDVKTNABALLTEMQNK 1545

QY 523 LVTIGDAAAPHINHIYAVILPESVQIG--RSAFRONGANNLIFMGSKVKTLGEMAFLSNR 580

Db 1546 ITEILNNNETTNE-----EKGNDIGPVRAAY-EEGLNNI-----NAATTTG-- 1585

QY 581 LEHLDLSEKQQLTEIPVQAFSDNALKEVLLPASLKTIREEAF-KKNHLKQL----- 630

Db 1586 ----DVTTAKDTAVQVQOLHANPVKK---PAGKKELDQAAADKKTKQTEQTPNASQOEIN 1638

QY 631 ----EVASALSHIAFNALDDNDGDQFQDNKV-----VVKT--HNSYALADGEHF-- 674

Db 1639 DAKQVDTQLQAKTN-VDQSTNEYVDNAVKEGKAKINAVKTFSEYKQDALAKLEDAYN 1697

QY 675 --IVDPD----KLSSTIVDL-EKILKLEGLDYSTLROTTQ-----TQFRDMT- 715

Db 1698 AKVNEADNSASTSSEIAEAKQKLAELKQTADQNVNQATSKDDIEVQIHNDLDNINDYTI 1757

QY 716 TAGKA-----LLSKSNLRQEKQKFL---QEAQFFLGRV-----D 747

Db 1758 PTGKESATDLYAYADQKKNISADTNTATQDEKQQAQKQVQNVQTALESINNGVDNGD 1817

QY 748 LDKATAKAEKALVTKKATKNGOLLERSINKAVLAYNNSAIKKANVKRLEKELDLTGT-- 804

Db 1818 VDDALTQCKAAL--DAIQVDAVVKPKANQAI-----EVKAEDTKESIDQSDQITAEK 1868

QY 805 -----LVEGKGPIAQAQTMVGQVYLLKTPPLPYIYIGL-----NVPFKSGK-- 846

Db 1869 TEALAMIKQITDQAKQIGITDATTAEVEKAKAQ-----GLEAFDNIQIDSTEKQKAI 1920

QY 847 --LIYALDM-----SDTIGEGOKDAYGNPILNVDEDNEGYHALAVATLADY-EGLDI 895

Db 1921 BELETALDQIEAGVNVNADATTE-EKEAFTNAL-----EDILSKATEDISDQTTNABI 1972

QY 896 KTIILNSKLSQTSIROVPTAAVYHRAGI-----FOAIONAAAEAE----- 934

Db 1973 ATVKNASLEQKAQRINPEVKNNALEAREVVNKKQIEIIKNAADADASAKEIARTOLGRYP 2032

QY 935 ----QLPKPGTHSEKS-----SSSESANSKDRGLQSNPKTNRGRHSAIL 975

Db 2033 DRFADKDKTQTNVAVELQNVTTIPAEIAIVPQNDPDANDTNGINDNDATANSNANATP 2092

QY 976 PRTG 979

Db 2093 ENTG 2096

RESULT 96  
ABU07438  
ID ABU07438 standard; protein; 3210 AA.  
XX  
AC ABU07438;



QY 955 SKDRGLQSNPKTNRGHRSAIL 975  
 Db 2979 TQVAHLCSQSQDSRGSPLL 2999

## RESULT 97

ADJ66483  
 ID ADJ66483 standard; protein; 3210 AA.

AC ADJ66483;

XX 06-MAY-2004 (first entry)

XX CenP-F kinetochore protein for anti-cancer protein complex.

XX neuroprotective; cytostatic; gene therapy; protein complex;

KW cellular network; cancer; neurodegenerative disease; drug target.

XX Homo sapiens.

XX WO2004009622-A2.

XX 29-JAN-2004.

PF 18-JUL-2003; 2003WO-EP007835.

XX 19-JUL-2002; 2002EP-00016109.

PR 19-JUL-2002; 2002EP-00016111.

PR 19-JUL-2002; 2002EP-00016123.

PR 19-JUL-2002; 2002EP-00016128.

PR 22-JUL-2002; 2002EP-00016427.

XX (CELL-) CELLZOME AG.

XX Merino A, Bouwmeester T, Bauer A, Drewes G, Marzioch M, Kruse U;

PI Superti-Furga G, Eberhard D, Ruffner H, Hobson S, Helftenbein G;

PI Cruciat C;

XX WPI; 2004-123372/12.

XX New protein complexes of cellular networks underlying the development of  
 PT cancer and other diseases, useful for diagnosing and/or treating  
 PT neurodegenerative diseases or cancer, and in drug screening.

XX Disclosure; SEQ ID NO 13; 809pp; English.

XX The invention relates to a protein complex of cellular networks  
 CC underlying the development of cancer and other diseases. The complex (i)  
 CC comprises at least one first and second proteins selected from any of the  
 CC proteins listed in the specification, or their functionally active  
 CC derivatives, fragments, homologues or variants, the variants being  
 CC encoded by a nucleic acid that hybridizes to the nucleic acid encoding  
 CC the protein under low stringency conditions. A complex (ii) comprises at  
 CC least two of the second proteins, where the low stringency conditions  
 CC comprise hybridization in a buffer comprising 3% formamide, 5 x SSC, 50  
 CC mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% BSA, 100 microg/ml  
 CC denatured salmon sperm DNA, and 10% (wt/vol) dextran sulfate for 18-20  
 CC hours at 40 deg C, washing in a buffer consisting of 2 x SSC, 25 mM Tris-  
 CC HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS for 1.5 hours at 55 deg C, and  
 CC washing in a buffer consisting of 2 x SSC, 25 mM Tris-HCl (pH 7.4), 5 mM  
 CC EDTA, and 0.1% SDS for 1.5 hours at 60 deg C. The composition and methods  
 CC are useful in diagnosing or treating diseases and disorders, preferably  
 CC neurodegenerative diseases. These may also be used as a drug target or in  
 CC manufacturing a medicament for the treatment or prevention of the above-  
 CC mentioned diseases or disorders. The composition may also be used for  
 CC treating cancer. This sequence represents one of the proteins of the  
 CC complex of the invention.

XX Sequence 3210 AA;

XX Query Match

XX Best Local Similarity 3.7%; Score 187.5; DB 8; Length 3210;

XX Matches 221; Conservative 179; Mismatches 400; Indels 301; Gaps 52;

QY 6 KTVALTTLTVSVVTHNQVFSLVKBPILKOTQOASSISGADYAESGSKLKNINSGPV 65  
 Db KTTALDQLSEKMEKTEQL-----ESHQSECLHICQVAEAEVKEKTELLQT--L 2115  
 QY 66 DDTVTDLFSKRTTPEKI-----KONLKGPREQELKAVTENTSEKQITSGSQBSQES 121  
 Db SSDVSELLKDKTHLQELQSLSEKDSQALSTKCELENQIAQLNKEKELL-----VKES 2168  
 QY 122 LSLNKTVPSTNWEICD-----FITKGNLTVLGS-----KSGVEKLS---QTD 161  
 Db ESLQARL--SESDYEKLVNSKALEAALVEKGFALRLSSTQEEVHOLRGIEKLRVRIEAD 2227  
 QY 162 -----HLV--LPSQAADGTQITQIVASFAFTPDKTAIAEYTSRACENGIEISOLDVDG--- 211  
 Db EKKQLHIAEKLKERENDSL-----KDKVENLERELQMSSENLVILDAENSKA 2278  
 QY 212 -----KEIINEGEVFNYSLLKKVTIPTGYKHI-----CO-----DAF 243  
 Db EVETLKTQIEEMARSLKVFE--LDLVTLRSEKENLTQIOBKQKQOLSELDKLLSFKSL 2335  
 QY 244 VDNKNIAEVNLPESLETISDYAFALHAKQIDLPNLKAIAGELAFDQITGKLSLPROL 303  
 Db LEEKEQAEIQIKESKTAVEM-----LQNLKEL-----NEAVALCGDOSI 2377  
 QY 304 MRLAERAFKSNHIKTIIEPRGNSLVIGEASFQDNLSQMLPDGLEKIESEAFNGPGDD 363  
 Db MKATEQSLDPPIEBEHQLR-NSIEKL-PARLEADEKKQLCVLQQLKESHH-----D 2428  
 QY 364 HYNNRV-----VLTWTKSGKNPGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKN 417  
 Db LLKGRVENLERELEIARTNQEHAALAEANS-----KGEVETLKAKTEGTMQ--- 2474  
 QY 418 SVTGFSNKGLQ-----KVRKNKLEIPKQHNGVTYITEIGDNAPRVDFQNKTRKYDL 470  
 Db -----SURGLELDVVVTIRSEKEDLTNELOKQERISELEIINSSPENT--LOEKEQEKVQM 2528  
 QY 471 BEVKLPSTIRKIGAFAPFOSNNLKSFEASDDLEEIK-EGAFMNNRITLLEKDKLVITGDA 529  
 Db KEKS--STAMEMLQTLQELNERNVAALHNDQACAKQONLSSQVECELEKQAQLQGLD 2586  
 QY 530 AFPHINHIYAVLPESV-----QEIGRSAPRQNGANNLI FMGSKVKTGLGMAFLSNRLEHLD 585  
 Db EAKQNY---IVLQSSVNGLIQEV-----EDGKQKL-----EKKDEBISRLKNQIQ--- 2628  
 QY 586 LSEQKQLTEIPQAFPSDNAL--KEVLLPASLKTIREAPFKNHLKQLEVASALSHTAFNAL 644  
 Db -----DQEQVLSKLSQVEGEHQLWKEQL--ELRNLTVLELQKIQLVQSKNAS-----L 2675  
 QY 645 DDNDGDEQFDNKVVVTKHNSVALADGBHFIVDPDKLSSTIVDLLEKILKLTIEGLDYSTLR 704  
 Db -----TLEVLQSSYKULENELEITKMDKMSF-----VEKVM-----TAK 2713  
 QY 705 QT-TQTQPRDMMTAGKALLSKNSRQGEKQKFLQBAQFFLGRVLDLDAIAKAERKALVTKK 763  
 Db ETELQREHMHM--AQKTAELQBEL--SGEKNRLAGELQLLLEI-----K 2754  
 QY 764 ATKNGQLLERSINKAVLAYNNSAIKKA-----NVRKLEKELDLTLGLVGKGLPLAQATVWQ 819  
 Db SSKD--QLAK-----LTLENSELKSLDCHMDQKQVEKGVREIBAEYQLRLEHAEKKH 2806  
 QY 820 GVIYLLKTEPLPPEYVIGLVNVDK-----SGKLIIVALDMSDTIG 858  
 Db QALLDNTN--KQVEVEIQTREKLTSEKCLSSQKLEIDLLKSKSEELNLSKATQILL 2863  
 QY 859 EQQKDA-----YGNPILNVDENEGYHALAVATLADYEGLDIKTILNSKLSQL----- 906  
 Db EELKTKMDNLKYNQNLKENERAGCKMKLLIKSCQKLE--EKEILQKELSQLAAQEK 2921  
 QY 907 -----TSIRQVPTAAYHRAGIFQAIQNAAAAEQLLPKPG-----THSEKSSSESAN 954  
 Db QKTGTVMDTKVDLTTTEIKE---LKETLEETKKEADEYLDKYCSLLISHLEKAKEMLE 2978

```

QY 955 SKDRGLQSNPKTNRGHRSAIL 975
Db 2979 TOVAHLCSQSQKDSRGSPLL 2999
RESULT 98
ID ABB63571
XX ABB63571 standard; protein; 1348 AA.
AC ABB63571;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 17505.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200117042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL07674.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Disclosure; SEQ ID NO 17505; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 1348 AA;
Query Match 3.7%; Score 187; DB 4; Length 1348;
Best Local Similarity 18.6%; Pred. No. 0.023;
Matches 219; Conservative 159; Mismatches 344; Indels 454; Gaps 55;
QY 92 REQELKAVTESEKQTSQSLSLN-KTVPSTSNWEICDFTKGNLVGLS 150
Db 11 RLQRELAKRSRQKQKTAKLPSSQATKSLKCNQAAPKTT-----ENTF-QGL 60
QY 151 KSGVKLSQTDHLV--LPSQAADGTLIQVASFATPDKTAIAFYTSRAGEN----- 201
Db 61 KLTIEELDLSYNLRIRRIPEKAFDGLK-----DSLNELRNLNGLNLPFST 109
QY 202 -----GRISQLDVQKEI--INEGEVFNYSLLKVTIPTGYKHIGQDAFVFNKRIAV- 252
Db 110 AELHVLKNRLDLSGNKIKLEEG-----LLK-----GCMDL-KEFYIDRNSLTSPV 156
QY 253 ----NLPSLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSLPQLMRLAE 308

```

RESULT 99  
 ABB77984  
 ID ABB77984 standard; protein; 1312 AA.  
 XX  
 AC ABB77984;  
 XX

DT 22-OCT-2002 (first entry)  
 XX Amino acid sequence of a yeast RAD50 protein.  
 DE  
 KW Nucleic acid integration; homologous recombination; telomeric region;  
 KW RAD50.  
 XX  
 OS Saccharomycetes cerevisiae.  
 XX  
 XX EP1217074-A1.  
 XX  
 XX 26-JUN-2002.  
 PD  
 XX  
 XX 22-DEC-2000; 2000EP-00204693.  
 PF  
 XX  
 XX 22-DEC-2000; 2000EP-00204693.  
 PR  
 XX  
 XX (UYLE-) RIJKSUNIV LEIDEN.  
 PA (BINA-) STICHTING BINAIR VECTOR SYSTEEM.  
 XX  
 XX Hooikaas PJJ, Van Attikum H, Bundock P;  
 XX WPI; 2002-550409/59.  
 DR  
 XX  
 XX Directing integration of nucleic acid of interest to a sub-telomeric  
 PT region in an eukaryote with preference for non-homologous recombination,  
 PT by steering an integration pathway towards homologous recombination.  
 XX  
 XX Disclosure; Fig 5; 63pp; English.  
 XX  
 CC The specification describes a method for directing integration of a  
 CC nucleic acid of interest to a pre-determined site, where the nucleic acid  
 CC has homology at or around the pre-determined site, in a eukaryote with a  
 CC preference for non-homologous recombination. The method comprises  
 CC steering an integration pathway towards homologous recombination. The  
 CC method is useful for directing integration of a nucleic acid of interest  
 CC to a subtelomeric and/or telomeric region in an eukaryote with a  
 CC preference for non-homologous recombination. The nucleic acid of interest  
 CC comprises an inactive gene to replace an active gene, or vice versa, is a  
 CC portion of a gene delivery vehicle, confers a desired property to the  
 CC eukaryotic cell, or encodes a therapeutic proteinaceous substance or a  
 CC substance conferring resistance for an antibiotic substance to a cell.  
 CC The method is useful for improving gene targeting efficiency. The method  
 CC is useful in the replacement of an active gene by an inactive gene, for  
 CC e.g. for the inactivation of genes controlling undesired side branches of  
 CC metabolic pathways, to increase the quality of bulk products such as  
 CC starch, or to increase the production of specific secondary metabolites  
 CC or to inhibit formation of unwanted metabolites, and also to inactivate  
 CC genes controlling senescence in fruits and flowers or that determine  
 CC flower pigments. The method is also useful for replacing an inactive gene  
 CC by an active gene. For e.g. the replacement of a defective p53 by an  
 CC intact p53. Many tumours acquire a mutation in p53 during their  
 CC development which renders it inactive and often correlates with a poor  
 CC response to cancer therapy. By replacing the defect p53 by an intact p53,  
 CC e.g. through gene therapy, conventional anti cancer therapy have better  
 CC changes of succeeding. The method is also useful for therapeutic  
 CC proteinaceous substance integration. A tumoricidal gene can be delivered  
 CC to a pre-determined site present only in e.g. proliferating cells, or  
 CC present only in tumour cells, e.g. to the site where a tumour antigen is  
 CC expressed form. ABB77984-86 represent RAD50 homologues. RAD50 is involved  
 CC in non-homologous recombination  
 XX  
 SQ Sequence 1312 AA;  
 Query Match 3.7%; Score 186.5; DB 5; Length 1312;  
 Best Local Similarity 19.0%; Pred. No. 0.024;  
 Matches 200; Conservative 174; Mismatches 362; Indels 315; Gaps 49;  
 QY 21 NQEVSLVKEPIKQTOASSISGADYAESGSKLKNINETSQPVDDTVDLPSDKRTPP 80  
 DB 361 NRNHLSSLEKAFQHKPQGLSINSDMAQVNHMSQFKAFISQDLTDTIDQFAKIQLKE 420  
 QY 81 EKIKD-----NLAKGPREQ-----ELKAVTENTESEKQITSGSQSKESLSLN 125

Db 421 TNLSDLIKSITVDSQNLVYNNKDRSKLIHDSLEAEKLSFKSLSTQDSLHLENLTKTY 480  
 QY 126 KTVPTSTNWEICDFITKGNLTVG-----LSKSGVEKLSQTDHLVLSQAQD----- 171  
 Db 481 K--EKLQSWESENIIPKLNQKIEEKNNEMIILENOIERFQ--DRIMKTNQQAADLVAKUGL 536  
 QY 172 -----GTOLIQVASPAFTPDKKTAIAEYTSRAGEGEISQ-----LDVDGKEI-INEGE 219  
 Db 537 IKKSINTKLDEL-----QKITEKLQNDLSRIQRFVPLTQEFQRADLEMDFKLFIN--- 586  
 QY 220 VNSYLLKKVITPTGYKH-----IQDAFVNDKNTAEV-----NLPSL 258  
 Db 587 -----MQKNIAIANNKQMHEDRRYTNALYNLNTIEKDLQDNQKSKERKVIQLSENLPDC 641  
 QY 259 ETISDY-----APAHLLAKOIDLPDNLKAJGELAFFNQITGKLSLPQLMR 305  
 Db 642 -TIDYNDVLESTELSYKTALENLKHQHTTLEFNKAL-EIARDS-----CCY 688  
 QY 306 LAERAFKSNHIKTIEFRGNSLKVIGEASFQ-----DNDLSQLMLPDGLEKIESEAFGNPG 361  
 Db 689 LCSRKFNESFKSLQ--ELKTKTDANFEKTLKDTVQNEKEYLSLALLEKHIIITLSI 746  
 QY 362 DDHVN-----RVVLWTKSGKNP-SGLATENTYVNPDKSLWOES--FEIDYTKKLEEDF 412  
 Db 747 NEKIDNSQKCLEKAKEETKTSKLDELEVDSTKLDEKELAESEIRPLIEKFTYLEKEL 806  
 QY 413 TVQKN-----SVTGFSNKGLO-----KVK 431  
 Db 807 KOLENSSVTISELSYNTSDEGIQTVDELDRQQRKMDSLRELKRTISDLQMEKDEKVR 866  
 QY 432 RNKNLEIPKQNGVITTEIGONAF--RNVDFOFKTLRYDYLEEVK-LPSTIRKIGAFAPQ 488  
 Db 867 ENSRMINLIKELTVSIESSLTQKQNDISRSKR-----ENINDISRKVELEARIIS 922  
 QY 489 SNNLKSFPASDDLEIKGAPMNNRIETLEKDKLVTTGDAAFHNHI---YAVLVPESV 545  
 Db 923 LKNKKD-BAQSVLDKVK-----NER--DIQVRNKQKTVAD---INRLIDRFQTIYNEV 970  
 QY 546 QEIGRSFQRCQANNLI FWSGKVKTLGEMAFSLNLEHLDLSEKQKLTETPQAFSDNAL 605  
 Db 971 D-----PEAKGFDEL-----QTTIKLEL--NKAQMLELKEQLDLKSNV-----NEE 1011  
 QY 606 KEVLLPASLKITREAFKKNHLKQLEVASALSHIAFNALDNDGDGEQFDNKVVVKTHNS 665  
 Db 1012 KRKLADSN---NEEKNLKNLELLELKSQLOHI-----ESE 1044  
 QY 666 YALADGEHFI VDPDKLSSTIVDLKILKILEGLDYSTLRQTQTQFQDMWT--AGKALLS 723  
 Db 1045 ISRLDVQNAEARDKYQ-----EBSRL-----RTRFEKLSSENAGKL--- 1082  
 QY 724 KSNLROGEKQKFLQEAQFGLGRVLDLKAIAEAKALVTKK-ATKNGQLLERSINKAVLAY 782  
 Db 1083 -GEMKQLQNDLSLTHQLRTDYKIEKNYHKEWELQTRSFVTDIDIVYSKALDSAIMKY 1141  
 QY 783 NNSATKKNANVRLEKEL-----DLLTGLVEGKGPLAQATMVQGVYLK 825  
 Db 1142 H--GLKMQDINRIIDELMKRTYSGTIDITKIRSDVSVSTVKGKSYNRYVMYK----- 1193  
 QY 826 TPLPLPYEYIGLVYFDKSGK-----LIYALDMSDTIGEG-QKDAYGNPILNVD 874  
 Db 1194 -----QDVELDMRGCSAQGVLASIIRLALSETFGANGCVIALDEPTTNLDE 1242  
 QY 875 DNEGHALAVATLADYEGDLIDITLNSKLSQ 905  
 Db 1243 EN-----IESLAK-----SLHNIINMRHQ 1262  
 RESULT 100  
 ABR53887  
 ID ABR53887 standard; protein; 1312 AA.  
 XX  
 AC ABR53887;



XX 20-JUN-2003 (first entry)  
 XX Protein sequence #SEQ ID 2639.  
 DE Multi-protein complex; eukaryote; drug target; diagnosis.  
 XX Saccharomyces cerevisiae.  
 OS  
 XX EP1258494-A1.  
 PN  
 XX 20-NOV-2002.  
 XX  
 XX 20-DEC-2001; 2001EP-00130253.  
 PF  
 XX 15-MAY-2001; 2001EP-00111774.  
 PR  
 XX (CELL-) CELLZOWE AG.  
 PA  
 XX Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;  
 PI Marzioch M, Schultz JD, Superti-Furga GD;  
 PI  
 XX WPI; 2003-250078/25.  
 DR N-PSDB; ACC61929.  
 DR  
 XX New isolated protein complexes useful for diagnosing a disease or  
 PT disorder, or as a target for an active agent of a pharmaceutical,  
 PT preferably a drug target in the treatment or prevention of disease or  
 PT disorder.  
 XX  
 XX Disclosure; SEQ ID NO 2639; 17pp + Sequence Listing; English.  
 PS  
 XX The invention relates to multiprotein complexes from eukaryotes. Proteins  
 CC of the invention and DNA sequences encoding them are given in records  
 CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are  
 CC obtainable by using a protein as a bait and isolating the set of proteins  
 CC which is attached thereto from cells. Such protein complexes may comprise  
 CC up to 30 distinct proteins. Protein complexes of the invention are useful  
 CC for diagnosing a disease or disorder, or as a target for an active agent  
 CC of a pharmaceutical, preferably a drug target in the treatment or  
 CC prevention of a disease or disorder. Note: The sequence data for this  
 CC patent is not represented in the printed specification, but is based on  
 CC sequence information supplied by the European Patent Office. The complete  
 CC document is available on CD-ROM  
 XX  
 SQ Sequence 1312 AA;

Query Match 3.7%; Score 186.5; DB 6; Length 1312;  
 Best Local Similarity 19.0%; Pred. No. 0.024;  
 Matches 200; Conservative 174; Mismatches 362; Indels 315; Gaps 49;

QY 21 NOEVFLVKEPILKQTSASSISGADYABSSGSKLKLKINETSGPVDVDTVTLDFSKRTTP 80  
 DB 361 NNNHLSLKEAFQHKFQGLSNTIENSDMAQVNHMSQKFAFISQDLTDTIDQPAKIQLE 420  
 QY 81 EKIKD-----NLAKPREQ-----ELKAVTENTSEKQITSGSQLEQSKESLSLN 125  
 DB 421 TNLSDLIKSITVDSONLEYNKDRSKLIHDSBELAEKLSFKSLSTQDSLNLHELENLKY 480  
 QY 126 KTVPTSNWEICDFTKGNLVG-----LSKSGVEKLSQTDHLVLPQAAAD----- 171  
 DB 481 K-EKLQSWESNIIPKLNQKEKNEMIIENQIEKQ--DRINKNQADLYAKGL 536  
 QY 172 -----GTQLIQVAFAPDKKTAIAEYTSRAGEISQ-----LDVDGKEI-INEGE 219  
 DB 537 IKKSINTKLDEL-----QKITEKLQDSRIRQVFPFQEFQADLEMDFOKLFIN--- 586  
 QY 220 VFNVSLLKKVTPTGYKH-----IGQDAFVDKNIAEV-----NLPESL 258  
 DB 587 -----MQKNIAINNNKMHBLDRRYTNALYNLTIEKQDQDNQKSKVKIQLSENLPEDC 641  
 QY 259 ETISDY-----AFAHLALKQIDLPDLNKLKAIQELAPFQNTGKLSLPRQLMR 305

DB 642 -TIDEYNDVLEETEISYKTALENLKMHTTLEPNKAL-EIAERDS-----CCY 688  
 QY 306 LAERAFKSNHIKTTFEFGNSLKVIGEASFQ-----DNDLSQMLPDGLEKIESEAFQNGP 361  
 DB 689 LCSRKFENESFKSLQ--ELKTKTDANFEKTLADTVQNEKEYLHSLRLEKHIITLNSI 746  
 QY 362 DDHYNN-----RVVLMTKSGKNP-SGLATENTVVPDKSLWQES--PEIDYTKWLEBDF 412  
 DB 747 NEKIDNSQKLEKAKBETKTSKLEDEVDSTKLEKELAESEIRPLIEKFTYLEKEL 806  
 QY 413 TYQKN-----SVTFGSNKGQ-----KVK 431  
 DB 807 KLENSSKTISEELSINYNTSEGIQTVDELQDQKWNDSRELKRTISDLQMEKDEKVR 866  
 QY 432 RNKNLEIPKQHGVTITEIGNAF--RNVPQNTKRYDLEEVK-IPSTIRKIGAFAPQ 488  
 DB 867 ENSRMNLIKELTVSEISSLQKQKIDDSIRSKR-----ENINDIDSRVKELEARIIS 922  
 QY 489 SNNLKSPEASDDLEIEKEGAFPMNRRIETLEKDKLVITIGDAAPHNHI---YAIVLPEV 545  
 DB 923 LKNKKD-EAQSVDKVK-----NER--DIQVRNKQKTVD---INRLIDRFQTIYNEV 970  
 QY 546 QEIGRSAPRONGANNLIFMGSKVKTGLGEMAFLSNRLHLDLSEQKQLEIPVQAFSDNAL 605  
 DB 971 D-----PEAKGFDEL-----QTTIKELEL--NKAQMLELKEQDLKSNV-----NEE 1011  
 QY 606 KEVLLPASLKTIRREAFKQNHLEKOLEVASALSHIAFNALDNDGDGDEQPDNKKVVKTHNS 665  
 DB 1012 KRKLADSN---NEEKNLQNLLELILKSLQHI-----BSE 1044  
 QY 666 YALADGEHFIVDPDKLSSTIVDLLEKILKLBGLDYSTLRQTQTPFRDMTT--AGKALLS 723  
 DB 1045 ISRLDVQNAEARDKYQ-----BESURL-----RTRFEKLSENAGKL--- 1082  
 QY 724 KSNLRQGEKQKFLQEAQFPLGRVLDKALAKAEKALVTKK-ATKNGQLLERSINKAVLAY 782  
 DB 1083 -GEMKQLQNOIDSLTHQLRTDYKDIENYHKEWELQTRSFVTDIDVYSKALDSAIMKY 1141  
 QY 783 NNSAIKXANVRLKEL-----DLTLGLVEGKGLAQATWVGQVYLLK 825  
 DB 1142 H--GLKMQDINRIIDELWKRTYSGTDITIKIRDSVSTVKGSYNYRVVYMK----- 1193  
 QY 826 TPLPLPEYIIGLVVYFQKSGK-----LIYALDMSDTIGEG-QKDAGNPLINVDE 874  
 DB 1194 -----QDVELDMRGCSAGQKVLASIIIRLALSETFGANGCVIALDEPTTLNDE 1242  
 QY 875 DNEGYPHALAVATLADYEGLDIKTILNSKLSQ 905  
 DB 1243 EN-----IESLAK-----SLHNIINMRHQ 1262

Search completed: August 28, 2005, 10:48:29

Job time : 229 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:37:35 ; Search time 48 Seconds  
(without alignments)  
2020.552 Million cell updates/sec

Title: US-10-078-531-2  
Perfect score: 5080  
Sequence: 1 MKKHLKTVALLTITVSVVTH.....LGYTSVALLSLITAIKKKKY 1008

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 110 summaries

Database :

PIR 79:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	331.5	6.5	1081	2 T31094	surface antigen Ba
2	221	4.4	690	2 F98114	choline-binding pr
3	219	4.3	1208	2 T39068	coiled coil protei
4	217.5	4.3	1072	2 A86827	hypothetical prote
5	214.5	4.2	3225	2 I52300	giantin - human
6	214.5	4.2	3259	1 A56539	Giantin - human
7	213.5	4.2	621	2 A95250	choline binding pr
8	212.5	4.2	3216	2 C90538	hypothetical prote
9	209.5	4.1	1227	2 C97033	uncharacterized pr
10	209.5	4.1	1252	2 B42771	reticulocyte-bind
11	208.5	4.1	2829	2 A42771	reticulocyte-bind
12	206.5	4.1	2269	2 T28677	rhoptry protein -
13	204.5	4.0	2481	2 D90011	FmtB protein (impo
14	204.5	4.0	4919	2 T31105	hypothetical prote
15	203.5	4.0	2346	2 T13829	Tpr homolog - frui
16	203	4.0	6713	2 B89921	hypothetical prote
17	199.5	3.9	2285	2 T12796	probable transglyc
18	199	3.9	2186	2 H89960	hypothetical prote
19	197.5	3.9	1269	2 F84730	probable myosin he
20	197.5	3.9	1790	2 S67593	transport protein
21	197	3.9	1119	2 H70126	surface-located me
22	197	3.9	1127	2 T28317	ORF MSV156 hypothe
23	197	3.9	2441	2 D71623	erythrocyte membra
24	196.5	3.9	1875	2 S38173	myosin-like protei
25	196.5	3.9	2139	2 T18296	myosin heavy chain
26	195.5	3.8	1778	2 AF1116	internalin protein
27	194.5	3.8	1108	2 T18353	protein P97 - Myco
28	194	3.8	949	2 E71940	translation initia
29	191	3.8	1365	2 T30822	lmpl protein - Myc

myosin-like coiled  
hypothetical prote  
protein T24H7.2 [i  
hypothetical coile  
probable peptidogl  
adhesin - Mycoplas  
RAD50 protein - ye  
surface exclusion  
chromosome segrega  
NMDA receptor-bind  
hypothetical prote  
translation initia  
hypothetical prote  
sp8 protein - firs  
probable cell surf  
hypothetical prote  
groovin gene prote  
major blood-stage  
adenylate cyclase  
hypothetical prote  
myosin heavy chain  
surface-located me  
conserved hypothet  
130K surface exclu  
restin - human  
probable secreted  
rhoptry protein -  
hypothetical prote  
internalin protein  
hypothetical prote  
dextranucrase (EC  
chromosome segrega  
hypothetical prote  
mhl protein - Myc  
probable peptidogl  
microtubule-vesicl  
hypothetical prote  
microtubule bindin  
RESA-H3 antigen PF  
F1003.10 protein -  
adenylate cyclase  
phage-related prot  
nuclear migration  
kinesin-related pr  
hypothetical prote  
probable transcrip  
uncharacterized pr  
hyaluronidase [imp  
hypothetical prote  
ATPase ScII, chrom  
hypothetical prote  
EF protein - Strept  
probable S-layer p  
nuclear/mitotic ap  
probable glutamate  
tanabin - African  
probable heat shoc  
IGA-specific metal  
myosin heavy chain  
protein F59B2.12 [i  
nucleotide-binding  
protein L precursor  
protein F21J9.12 [i  
myosin heavy chain  
P115 homolog - Met  
zinc metalloprotei  
probable myosin he  
gp150 protein - fr  
myosin heavy chain  
hypothetical prote  
hypothetical prote



```

QY      HALVAVALDLEGGDIKIENSNTSQUISIKUVFIMAFHRAUGFQALQANNAEQUELLEFR 353
880
Db      1024 --LMTSNVTDSQMLKLTREALQSQTNKNIDHLSTILERNKEYKSLDDYNNQLRARYKN 1081

```

QY P0THSESSSSANSKDLGQSNPK-----INNGRHSAILLPFGSKSGFVIGLGLISVA 395  
940 ::::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db 1082 LQSNTPOSTSQGYESEIKGLSKLTLYLQSKCRREHSLRLDIAFSKFILMQLTGTCN 1141

QY	996	LIST	999
	:	:	:
Db	1142	KINL	1145

RESULT 4  
A86827  
hygrotherical protein vafg [imported] - Lactococcus lactis subsp. lactis (strain

C:species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: A86827  
R:Botolin, A.; Winkler, P.; Mauger, S.; Jaillon, O.; Maitarme, K.; Weissenbach,  
B.

Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium *Lactococcus*  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: A86827

A:status: preliminary  
A:Molecule type: DNA  
A:Molecule type: 1-1072 <STO>  
A:Residues: 1-1072  
A:Cross-references: UNIPROT:O9CF64; GB:AE005176; PID:q12724625; PIDN:AAK05715.1

A:Experimental source: Strain 101403  
C:Genetics:  
A:Gene: yqfg

Query Matchn 4.3%; Score 217.5; DB 2; length 1072;  
Best Local Similarity 16.9%; Pred. No. 0.02;  
Matches 151; Conservative 159; Mismatches 365; Indels 219; Gaps 29;

```

QY 149 LSKSGVSKSI-----DLRVVFSQADG|-----QLQVAGSFAPFDK 189
      : : : : : | : : : : : |
Db 1 MSQSIKKITWTGVGLLTATLNPFTLNLHQLLTPEVATTSKITYTDEQKITNLDKVS 60

```

QY	ALAEITSRAGENG-EISQLDWDKNEINEGE-----VPSILNKNLIFIGIANH	238
D6		
D7		
D8		
D9		
D10		
D11		
D12		
D13		
D14		
D15		
D16		
D17		
D18		
D19		
D20		
D21		
D22		
D23		
D24		
D25		
D26		
D27		
D28		
D29		
D30		
D31		
D32		
D33		
D34		
D35		
D36		
D37		
D38		
D39		
D40		
D41		
D42		
D43		
D44		
D45		
D46		
D47		
D48		
D49		
D50		
D51		
D52		
D53		
D54		
D55		
D56		
D57		
D58		
D59		
D60		
D61	LTAEVASYAGLAGADIVIPDI-----IYNNGQTAYAITSIGTYAFSNSGIRSVIIGNVVVDI	116

QY -----NANABVNPSSEIISDIAPAHALAAQIDIDPUNALGELAFD 239  
GQAFVU-----:::--:::--:::--:::--:::--:::--:::--:::--:::--:::--  
DB 117 NTSAFOTTAFDDYKKSLLTKVVLGAKVQNIKTDAFAGNAISSIEFPNSVLKATRAFAN 176

QY : : : : :  
291 NLLIGKSLFQMKTABRAPSNNHINLIEFRGNSLVIVGEASFOUNDSQLTFFDGJER 350  
DB : : : : :  
177 NNLT-ELSLGNITEIMAKAFOSNQITTEFADESILTVDSAAGSGSVQSUTLGIGV-- 233

QY	1E5EAF1GNFGDHRNNKVVLMNSGA-----NFSGUAERINIVNDRSLWJES	359
	: : : : :         : : : : :	
Db	234 -----TLADDFVNKTSPLEQLSDLPDEIRTSVNSSGL-FDKSWINDSS---QS	281

```
QY      PDLDITRWLEEDFPIQANSVIGFSNAGLQRKVRKNLNLSIPAQHNGVV-----ELGJN 453
          :::::|
          :|||:|
          :|||:|
          :|||:|
DB      -----TENADTTPVSSNSVEQVDVASS-ESTDANSAASLYPISEASSVTDN 331
```

```
QY 454 AFRNVDFONKTLKYLDEVLKPLSTIRKIGAFAPFOSNLSKFEASD-----DLEE 503
Db 332 TLNLSISLSDSSISQSTNSQ--SGASSTABISYDSSENSQNSQNSNSSEKDSNQ 389
QY 504 IKEGAFMNRRTETLEKDKLVITIGDAAPIHNIHVAIVLP--ESVQEIAGRAGFRQNGANN 560
Db 390 SSLGSSMSNBSHNSNSNINETNNSSEITNLPSPNPTESNVSVDQTSSEASTNSNS 449
QY 561 LIFMGSKVKTIGEMAFNLRLHLDLSEKQKLTETPVOAFSDNALKEVLLPASLKTRIEE 620
Db 450 ISLSPSNISSTSDSESATNSQDFSNVAEVANNSLASVNSSSVLSSTSTADNL--GINQS 508
QY 621 AFKNHKLQLEVASALSHIAFNALDDNDGDEQFNDKVVVKTTHNSYALADGEHFIVDPDK 680
Db 509 GSDNLTQDSSEISTSGAFLSNQTSS-----ASTNSNS-----SISLSPN 550
QY 681 LSSITVDLEKILKLEGLDYSLRTQTQTFRDMTTAGKALLSKSNLRQGEKQKPLQEAQ 740
Db 551 ISSISV---LESTSSSNFNVBAEVANNSLASVNSSSVLSSTSTADN-----LEINQ 601
QY 741 PFLGRVLDKAIKAEKALVTKATKN-----GQLLERINKAVLAYNNSAIKKNVYK 793
Db 602 FGSDNLTQDSSEISTSGAFLSNQTSSSEASSNSMSSINSPLSLSLTSNSSESATNSQ-- 659
QY 794 RLEKELDLTLGLVEGKGPLAQATWQGVYLLKTPLPPEYVIGLVNVPDKSGKLIYALDM 853
Db 660 -----SSEATKV-----DNNNS-----STHS 674
QY 854 SDTIGEGQKAVGNPILNVDENEGYHALAVATLADYEGDLDTILNSKLSQTSIROVP 913
Db 675 SNILSGNDSGDS--DSDSDSDS-----SNLSSSPNLETNQTISKPEVNINISNP 725
QY 914 TAAVHRAGIFQAIONAAABEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTN 967
Db 726 KKV-----SSNSVQENSTDMETNPKSS 750

RESULT 5
152300
gi|152300|ref|NM_001001001.1|giantin - human
N:Alternate names: gcp372
C:Species: Homo sapiens (man)
C>Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 09-Jul-2004
C:Accession: I52300
R:Sonoda, M.; Misumi, Y.; Fujiwara, T.; Nishioaka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in
A:Reference number: I52300; MUID:95100974; PMID:7802676
A:Accession: I52300
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RES>
A:Cross-references: UNIPROT:Q14789; GB:D25542; NID:5662389; PIDN:BAA05025.1; PID:g808869
C:Superfamily: giantin

Query Match 4.2%; Score 214.5; DB 2; Length 3225;
Beat Local Similarity 19.8%; Pred. No. 0.14;
Matches 212; Conservative 168; Mismatches 427; Indels 263; Gaps 45;

QY 20 HNOVFSLVKPEILKQATQASSISGADYAEBSGKSKLKI-----NETSGPVDVDTDL 72
Db 199 HEDELQLV-----TQAD-----VETEMQOKLRVLRQKUBEHESLVGRAQVVDL 243
QY 73 FSDKETTPEKIKNDLAKGPREQLK-----AVTENTESEKQITSGSQLEQSKESLSLNK 126
Db 244 LQOELTAAEQRNQILSQLOQMEAEHNTLVETEREESKLLKLEKMELEVAERKLSFH- 302
QY 127 TVPSTNSNWEICDFTYKGNLVL-----SKSGVEKLSQTDHLVLPQADGTLQIQV 178
Db 303 NLQEMHLLLEQFEQAGQAQAELESRYSALEQKHAEKMEETSHLSLQKTG--QELQS 359
QY 179 ASFAFTPKKTAIAEYTSRAGENGRI-----SQLDVGDKGKIIINEGEVFNLSYLLKKVTIPT 233
```

```
Db 360 ACDALKDQNSKLLQDKNEQAVQSQTIIQOLEDQLOQSKSEI-----SOFLNRLPQQ 411
QY 234 GYKHIGQDAFVDNRKNIAEAVNLPESLETISDYAFAPHLAKQIDLDPNLKAIGELAFQDNQI 293
Db 412 -HETASQTSPPDVN-----EGQAVTEENIASLQKRVVEL-----ENEK 450
QY 294 TGKLSLPRQLMRL--AERAFKSNHKTIEFRGNSLKVIGEASFQDNDSQLMLPQGLEKIE 352
Db 451 GALLSSIELEELKAEKELSSQITLLEAQNR-----GEA---DREVSEISIVDIANKRS 503
QY 353 SEAFTEGPGDDHNNRVVLTWKSGKNTSGLATENTYNNPDKSLWQSEPEIDYTKWLEEDF 412
Db 504 SSA-----EESQD-----VLENTFSQKHKL-----SVLLEMEKAEQBEI 539
QY 413 TYQKNSVTG-----FSNKGLOKVRKNKYLE-----IPKQHGVT 446
Db 540 AFLKLQLOQKRAEADHEVLQDKEMQMBEGEAPIAKMKVFLDTGQDFPLMPNEESSLP 599
QY 447 ITEIGDNAPRVDFONKTLRKYDLEE--VKLPSTIRKIGAFAPFOSNLSKFEASDDLEEI 504
Db 600 AVE---KEQASTEHQSTSEISLNDAGVELKST-----KQDGDKLSAVPDTGQC 647
QY 505 KEGAFMNRRTETLEKDKLVITIGDAAPIHNIHVAIVLPESVQEIQR-----SAFRONGA 558
Db 648 HQDELERLKSQILEL-----ELNFHKAQEIYKELNDEKAKEISNLNLQIEFFKNAD 699
QY 559 NN-----LIFMGSKVKTIGEMAFNLSNRLEHLDL-----SBQKQLTETIPVQAFS 601
Db 700 NNSAFTALSEERDQLL---SQVKELSMVTELRAQVQKLEMLNLAEEAERQRLDYESTAH 756
QY 602 DNALKVEL---LPASIKTIREBAFKQKHLKQLEVASALSHIAFNALDDNDGDEQFQDNK- 656
Db 757 DNLLTEQIHSLSIEAKSKDKVIEVL-QNELDDVQLQFSEQSTLIRSL-----QSLOQNK 810
QY 657 -----VVVKTHNSYALADGEHFIVDPDKLSSSTIVDLEKILKLEGLDYSLR 704
Db 811 SEVLEGAERVHISSKVEELSQALESKELETKMDQLL-----LEK-----KDDVETLQ 859
QY 705 QTTQTPQFRDMTTAGKALLSKSNLRQGEKQKPLQEAQFPLGRVLDKAIKAEKALVTKKA 764
Db 860 QTIEEKQQVT-----EISPSMTEKVMQVQNEEKFSLG-VET-KTLKEQLNLSRAEE 909
QY 765 TKNQOLLERSINKAVLAYNNSAIKKA---NVKRLKELDLTLGLVEGKGPLAQATWQGV 821
Db 910 AKCEQVEEDNEVSSGLKQNYDEMSPAGQISKEELQHEFDLLKCKENQKRLQALINRK 969
QY 822 YLLKTPLPPEYVIGLVNVPDKSKLIYALDMSDT-IGEGOKDAYGNPILNVDNDEGYH 880
Db 970 ELLQVSRLEBELANLK---DESKK---EIPLSSTERGEVEE-----DKNKEYS 1013
QY 881 ALAVATLADYEGDLTKILNSKLSQTSIR-----QVPTAAVHRAGI 922
Db 1014 EKVTSKQETIEIVLKQITISEKEVELQHIRKDLBEKLAABEQFQALVKQNMQLQDKTNQ 1073
QY 923 FOATQNAABAEQQLPKPGTHSEKSSSES-ANSKDRGLQSNPKTNRGH 971
Db 1074 IDLLQAEISENQAIQKLIITSNTDASDQDSVALVKETVVISPPCTGSSEH 1123

RESULT 6
A56539
gi|A56539|ref|NM_001001001.1|giantin - human.
N:Alternate names: macrogolin
C:Species: Homo sapiens (man)
C>Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 09-Jul-2004
C:Accession: A56539; S37536
R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994
A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein (
A:Reference number: A56539; MUID:94187728; PMID:7511208
A:Accession: A56539
A:Molecule type: mRNA
```

A:Residues: 1-3259 <SEE>  
A:Cross-references: UNIPROT:Q14789; EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405714  
C:Gene: GDB:GOLGB1; GCP: GCP371  
A:Cross-references: GDB:454958  
A:Map position: 3q13.31-3q13.31  
C:Superfamily: giantin  
F:3238-3254/Domain: transmembrane #status predicted <TM>  
  
Query Match 4.2%; Score 214.5; DB 1; Length 3259;  
Best Local Similarity 19.8%; Pred. No. 0.14;  
Matches 212; Conservative 168; Mismatches 427; Indels 263; Gaps 45;  
  
QY 20 HMQEVSILKEPILKOTQASSISGADYAESSGSKLKI-----NETSGPVDDTVTDL 72  
DB 233 HEDELQLV-----TQAD-----VETENQKLRVLQKLEHEESLVGRAQVDL 277  
  
QY 73 FSDKRTPEKIKDNLAGPREQELK-----AVTENSEKQITSGSQLEQSKESLSLNK 126  
DB 278 LQOELTAABQRNQLISQQLQOMEAEHTLRNLTVETERESKILLEKWELEVAERKLSPH- 336  
  
QY 127 TVPSTSNWEICDPIKGNLVL-----SKSGVEKLSQTDHLVLPQAADGTQLIQV 178  
DB 337 NQOEMHLLLEQFEQAGQAQAELESRYSALEQKHAEMEEKTSHLSLQKTG---QELQS 393  
  
QY 179 ASFAFTPKKTAIAEYTSRAGNGEI-----SOLDVDGKEIINEGEVFNFSYLLKKVTIPT 233  
DB 394 ACDALKQDNKLLQDNQEQVQSAQTIQOLEDQLOKSKEI-----SQFLNRLPIQQ 445  
  
QY 234 GYRHGQDAPVNDKNAEVLNLPESLETISDYAPAHALKQIDLPNMLKAIGELAPFDNQI 293  
DB 446 -HBTASQTSFPDYN-----EGTQAVTEENIASLQKRWEL-----ENЕК 484  
  
QY 294 TGKLSLPROLMRL-ARRAFKSNHIKTIIFRGNLSKIVIGASFDQNDLSQMLPDGLEKIE 352  
DB 485 GALLASSIELEELKAENKLSQITLLEAQNR-----GEA---DREVSEISIVDIANKRS 537  
  
QY 353 SEAFNTGNPGDDHYNRRVLTWKGKNGPSGLATENTVVPDKSLWQSPEDIDYTKLEEDF 412  
DB 538 SSA-----ERSGQD-----VLENTFSQHKEL-----SVLLENKKAQEBI 573  
  
QY 413 TYQRNSVTG-----FSNKGGLQKVRKNLE-----IPKHNGYV 446  
DB 574 AFLKLQLOKRAEADHEVLQDKEMQMEGEGIAPIKMKVFLDQTQDFPLMPNEESSLP 633  
  
QY 447 IYEGIDNAPRVDFQNTLRKYDLER--VKLPSTTRKIGAPFQSNLKSFAASDDLEI 504  
DB 634 AVE---KEQASTEHQSRTESEISLNDAGVELKST-----KQDGKSLSAVDPDIGC 681  
  
QY 505 KEGAFMNNRIETLELKDVLVTIGDAAPH-INHYIATVLPESVQEIQR-----SAPRNGA 558  
DB 682 HQDELERLSQILEL-----ELNFHQAQEIYKNLDEKAKETISNMLNQLLEEFKQAD 733  
  
QY 559 NN-----LIFMGSKVKTIGEMAFLSNRLEHLDL-----SEQQLTEIPVQAFS 601  
DB 734 NNSSAFTALSEERDQLL---SQVKELSMWTELRAQVKQLEMLAEARQRRLDYESQTAH 790  
  
QY 602 DNALKEVL-----LPASLKTIRBEAFKQNHKQLEVASLSHAFNALDNDGDGPDNK- 656  
DB 791 DNLTLTEQIHSLSIEAKSKQVKIEVL-QNELDDVQLQFSEQSTLIRSL-----QSOLQNKE 844  
  
QY 657 -----VVVKTHNSVALADGEHFIVDPDKLSSTIVDLEKILKLEGLDYSTLR 704  
DB 845 SEVLEGAERVHISSEKVBELSQALSQKELEITMDQLL-----LEK-----KRDVETIQ 893  
  
QY 705 QTQTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLQKAIKAEKALVTKA 764  
DB 894 QTIEEKDQVTT-----EISFSMTEKVVQLNEEKFSLG-VET-KTLKEQLNLSRAEE 943  
  
QY 765 TNGQLLERSINKAVLAYNNSAIKKA---NVKREKELDLGLVEGKGPLAQATWQGV 821  
DB 944 AKKEQVEEDNEVSSGLKQNYDEMSPAQIISKBELOHEFDLLKKENEQRKRLQAALINRK 1003

QY 822 YLLKTPLPLPEYYIGLVNVFQSKGLIYALDMSDT-IGEGQKDAYGNPILNVDNEGYH 880  
DB 1004 ELLQVRSLREELANLK---DESKK---EIPLSETERGEVEE-----DKNKEYS 1047  
  
QY 881 ALAVATLADYEGLDIKTILNSKLSQTSIR-----QVPTAAAYHRAGI 922  
DB 1048 EKCVTSTKQEIYILYKQITISEKEVLEQHIRKOLEEKLAAEQFQALVKQMNQTLQDKTKQ 1107  
  
QY 923 FQAIQNAARAEQLLPKPTHSEKSSSES-ANSKDRGLOSNNPKTNRGPH 971  
DB 1108 IDLLQAEISENQAIIOKLIITNTDASDGSVALVKETVVISPPCTGSSSH 1157  
  
RESULT 7  
A:Species: Streptococcus pneumoniae (strain TIGR4)  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: A95250  
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidon, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, neon, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: A95250  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-621 <KUR>  
A:Cross-references: UNIPROT:Q97NB5; GB:AE005672; PIDN:AAK76194.1; PID:g14973649; GSPDB:C  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP2136  
  
Query Match 4.2%; Score 213.5; DB 2; Length 621;  
Best Local Similarity 23.2%; Pred. No. 0.013;  
Matches 132; Conservative 75; Mismatches 175; Indels 187; Gaps 31;  
  
QY 6 KTVALTLLTVSVV-----THNQEVF-----SLVKEPILK-----Q 35  
DB 3 KTTILSLTTAAVILAAVYVNEPILADTPSSEVIEKTVGSIITQNNIKYKVLTVEGNIGT 62  
  
QY 36 TOASSISGADY-AESSGK---SKLKINETSGPVDDTVTDLS-----DK 76  
DB 63 VQVGNVTPVEFAGQDGRPTPIKTVGDKVFTVEVASQAFSYPDGTGRIVYPPSS 122  
  
QY 77 RTTPKIKDNLAGPREQELKAVTENTSEKQITSGSQLEQSK----- 120  
DB 123 ITIPSSIKKIQQKGFHSGKAKTII-----FDKGSQLEKIEDRAFDFSELEEELPAS 174  
  
QY 121 -----SLSLNK--TVPSTSNWEICDPIKGNLTVGLSKSGVEKLSQTDHLVLPQ 168  
DB 175 LEYIGTSAPFSQKLKLTFSSSSKLEL-----ISHEAFANLSNLEKLTLPKS 222  
  
QY 169 AAD-GTQLIQVA-----SFA-----FTPDKTAIAEYTSRAGENGISQLD 208  
DB 223 VKTLGSNLFRLTSLTKHVDVEEGNESFASVGVLFPSK-QTQLIYYPQ--KNDSYKTP 279  
  
QY 209 VDGKEIINEGEVFNFSYLLKKVTTIPTGYKHIGQDAFVNDKNIAEVLNLPESLETISDYAF-A 267  
DB 280 KETKELASYSFNKNSY-LKLELNEGLEKIGTFAPADAKLEEISLPSNLETIERLAFVG 338  
  
QY 268 HLAHQIDLPNMLKAIGELAF-----FDNQITKLSLPROMLAEARAFKSNHIK 317  
DB 339 NLELKEILIPDNVKNFGKVMNGLPKLSLTGNNIN---SLPSFSLGVLDSLKEIHIK 395  
  
QY 318 -----TIERGNS--LKVICEASFQ-DNDLSQMLPDGLEKIESEFT 357  
DB 396 NKSTFESVKQDTFAIPETVKFVYSEHTKDVKLSNLSNDI-----IVEKVDNFKQETDV 451  
  
QY 358 GNPQDDHYNRRVLTWK-----SGKNPSGLATENTVVPDKSLW---QESPEIDY 404



Db 452 AKP-KKNSQGVGVGVKDKGLWYLYNESGSMATGHV-----KDKGLWYLYNESGSM-A 502  
QY 405 TKWLEED--FTYQKNS---VTGF-SNKGL 427  
Db 503 TGWVKDKGLWYLYNESGSMATGWVKDKGL 531

RESULT 8  
C90538  
hypothetical protein MYPV 2110 [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
C:Species: Mycoplasma pulmonis  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C:Accession: C90538  
R:Chambaud, I.; Hellis, R.; Ferris, S.; Barbe, V.; Samson, D.; Galieson, F.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001  
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
A:Reference number: A99512; MUID:21267165; PMID:11353084  
A:Accession: C90538  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3216 <KUR>  
A:Cross-references: UNIPROT:Q98029; GB:AL445566; PID:g14089624; PIDN:CAC13384.1; GSPDB:C  
A:Experimental source: strain UAB CTIP  
C:Genetics:  
A:Gene: MYPV 2110  
A:Genetic code: SQC3

Query Match 4.2%; Score 212.5; DB 2; Length 3216;  
Best Local Similarity 19.9%; Pred. No. 0.17;  
Matches 218; Conservative 148; Mismatches 380; Indels 347; Gaps 55;

QY 7 TVALPLTT--VSVVTHNQEVPSLVK----EPILKQTOASSISGADYAESGSKLKN 59  
Db 2274 TLITITISSDOTSRITYQIEIGFVSDNRAENIVAKGYLSIEIOASGRAPESGTVIQT- 2332

QY 60 ETSQGVDP-----TVTDFDKRTTPB--KIKONLAKGPREQELKAVTENTSEKQITS 111  
Db 2333 ----PILDPVQNNSTDEFNKILTNHESAGLINKNPQ-----SNNSNLEQSILS 2381

QY 112 QSGL----BQSKESSLNKTVPSTSN-----WEI-CDFITK-GNTLVGLSKSGVEK 156  
Db 2382 AEVLTTNEVPAESVKLIQVSKKGDQDDHYSREYVWVYSGFIQRTGADILNIRRAIEAS 2441

QY 157 LSQTP----HLV-----LPSQADGTQLIQVASFAPDPK-KTAIAEYTSRAGE 200  
Db 2442 LRENDKAQHIVKVESREKITKIESTTTVGKNFFTTSDFEINNELIKTKVTNVQNDY 2501

QY 201 NGE1-SQLDVGKEIINEGEVPSYLLAKKVITPTQYKHIGQDAFVDN---KNIAFVNLPB 256  
Db 2502 NVELLSVLDSIRENIIISGNARVKIRSK-----DNSSEKSEFIVNI-S 2544

QY 257 SLETISDYAFALHALKQIDLDP-----NLKAIAGELAFF---DNQITGKLSLPRQL 303  
Db 2545 GFETLSEAFQYKIKSSTRFIPSVKANKDLWDTKATSENFPSQNKDNKFPD----- 2600

QY 304 MRLAER-----APKSNHIKTIETPRNSLKVIGASFDQNDLSQMLPDPGLEKTES 353  
Db 2601 -AVKERFVFSVDTSKVKFKMELSKTIKTNFGIESTFFQSAQVEEDV-----DGFKK--- 2650

QY 354 EAFQGNPGDDHNNRVVLWTKSGKNPSGLATENTVNPDKSLWQSPRIDYTKMLEEDFT 413  
Db 2651 ELFDN---DIFR---VQTEKDGK-----TINKYVKEKE---KNSHNIDFTK----- 2688

QY 414 YQKNSVTGFSNKGLOKVRKNKNLEIPKHONGVTIITEIGDNAPRNVDFQNKTLRKYDLEEV 473  
Db 2689 -----IKKLEI-----GNFSEI PKDYFKDAS-----SLTEL 2715

QY 474 KLPSTIRKIGAPAFOSNNLKSFEASDDLEEBEIKGAFPMNRITELBKLVITIGDAAPHI 533  
Db 2716 KINSQVSKIKESAFESAKLTSLELPSNLVEIGPNAPKNSVLTSL----- 2759

QY 534 NHIYAILVPESVQIEIGRSFRQNGANNLIF--MGSKVKTGLGEMAFLSNRLEHLDLSEQKQ 591

Db 2760 -----SGLETKLSALKEN-----VFEKANDSIKTT-----IWNWKEALKAKEPI 2800  
QY 592 LTEIPVOAFSDNALKEVLLPASLTKTIREAPKQNHKLKOLEVASALSHIAPNALDDNDGDE 651  
Db 2801 LT-----STNK-----IPSSD-----KKAKKPSQFEQ-----SNFSFAATKQEDSDI 2838

QY 652 QPD-----NKVVVKTTHNSYALADGEHFIVDPDKLSSTIVDL-----EK 690  
Db 2839 HFEITGFSQDDVFGTIITIKIKIKTKTDESFPDEN-----NPSK---KVLTFDKTNLEEK 2892

QY 691 ILKILEGLDY-----STLRQTTOTQPRDMWTACK---ALLSKSNLRQGEKQKFLQEAQF 741  
Db 2893 VTKIKDASSVPDSKIEITIEKKRDTSSNNTSTENQNKYRLVQWQKDVNGSGSVFLPOSSV 2952

QY 742 ---FLGRVDL-DKAIKAEKALVTKATKNGQLLE--RSINKAVLAYNN----- 784  
Db 2953 NAPPVETQIGDKTTSVLDFNPNPQSQIRNNSLFENLKPSSAYLVMSNDASLKNQSKIT 3012

QY 785 --SAIKANVKRLEKELDLTLGLVEGKPL-----AQNTWQGVYLL 824  
Db 3013 LLSTFKNSSTHSLHFESKVLGSGWIPGSGSLRFDQNNNIDRESQDVNKANIESTENNITA 3072

QY 825 KTLPLPLEYIG-----LVVYFDKSGKL-----IYAL 851  
Db 3073 KGSMP-QSFIQGESKTIILEYRDEQGRYEFYFFVFNNDKVKRYKVMKTEAVAKAYLI 3131

QY 852 DMS---DTIGEGQKDYGNPILNVDEDEGYPHALAVATLAD-----YEGLDIKTILNSK 902  
Db 3132 DLSSIGDT-GNSGFQAKLSVISPNQNNFLTYEERIKLLDELGKKWFNNISIKDSSGK 3190

QY 903 LSQLTSTRQVPTA 915  
Db 3191 VENNTYNNMKTA 3203

RESULT 9  
C97033  
uncharacterized protein, probably surface-located [imported] - Clostridium acetobutylicu  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: C97033  
R:Nolling, J.; Berton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: C97033  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1227 <KUR>  
A:Cross-references: UNIPROT:Q97K41; GB:AE001437; PIDN:AAK79054.1; PID:g15023995; GSPDB:G  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC1080

Query Match 4.1%; Score 209.5; DB 2; Length 1227;  
Best Local Similarity 18.1%; Pred. No. 0.056;  
Matches 237; Conservative 207; Mismatches 402; Indels 463; Gaps 68;

QY 1 MKKHLKTVALTTLTVSVVTHNQEVPSLVKPEILKQTOASSISGADYAESGSKKLKLINE 60  
Db 4 IKNTLLVFSISVTTSAV-----LLSKPAIKAYAADNSV-----YKSSISNSN-EINV 49

QY 61 TSGPVDVTVDLDFDKRTTPEKIKDNLAKGPREQELKAVTENTSEKQITSGSQLEOSKE 120  
Db 50 KGKEV-----QEKEVNVKNNSKNSKVSSNE-NQVSNKNSNPKVSSSSSEIQSINK 100

QY 121 SLSL-----NKTVPSTGNWEICDEITKGT-----LVGLSKSGVEKLSQTDHLVLP 166  
Db 101 NVNLQVQNNKSVLAASN--VDDVKKINSQVTSQDYIAIGETKVKPPTDLDIINKAIVD 158

QY 167 SQADGTQLIQVASFAPDPKTKTAIAEYTSRAGEN-GEISQLDVGKEINEGEV--FNS 223



QY 590 KQLTIPVQAFSDNALKEVLLPASIKTIREBAFK-KNHL-----KQLEVASALS 637  
 DB 731 ----YVGKITPELALTLLGDAKLTAQELKFSKNNVLTETNNKNTNELDVHKNQ 786  
 QY 638 ---HIAFNALDND-----GDS-----QFDNKV-VWTHNSYALA 669  
 DB 787 DAYKALREILAHSDRIDTKQKSSKLEIENGNIYIKVVLINQYKNKISSIKSKEBAVSUK 846  
 QY 670 DG-----EHP-----IVDPKLSSTIVDLKILKLEGLDYSLRQTT-----QTQ 710  
 DB 847 IGWVKKHSLSKITCSKSDYNIILAEKOTEL-QNLNRSPTQKNTNNSDKLEKIKTD 905  
 QY 711 FRDMTAGKALLSKN-----LRQEKQKFLQBAQFFLGRVDL 748  
 DB 906 FESLNALKTLGEVNALKASDNNHVSQSEPNPALSEIEK-----BETDIDLNTAL 961  
 QY 749 DKAIKAE-----KALVTKATKNGQL-----LERSINKAVLAYNNSAIKKNVKRL 795  
 DB 962 DELLKKGRTCEVSRVKLIKDTVTKEISDDTELINTIERKV-KAYLAY-----IKQYEDTV 1016  
 QY 796 EKELDLLTGLVEGKGLAQATMVQVYLLKTLPLPEYIYGLNV-----YFDKSGK--- 846  
 DB 1017 Q---DVL-----LNEHFNTKQVSNHPTNFKSNKSSE 1047  
 QY 847 -LIYALDMSDTIGEQKDAYGNPILNVDNEDNGYHALAVATLADYEGDLIKTILNLSKQ 905  
 DB 1048 ELTKAVTDSKTIISLKG-----IIEVNTE-----MNTIESSAKEIEALYNELKNK 1096  
 QY 906 LTSIRQV 912  
 DB 1097 KTSINEI 1103

## RESULT 11

A42771  
 raticulocyte-binding protein 1 - Plasmodium vivax  
 C1Species: Plasmodium vivax  
 C1Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 28-Apr-1995  
 C1Accession: A42771  
 R1Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.  
 Cell 69, 1213-1226, 1992  
 A1Title: A raticulocyte-binding protein complex of Plasmodium vivax merozoites.  
 A1Reference number: A42771; PMID:92315338; PMID:1617731  
 A1Accession: A42771  
 A1Status: preliminary  
 A1Molecule type: DNA  
 A1Residues: 1-2829 <GAL>  
 A1Experimental source: Belem strain, merozoites  
 A1Note: sequence extracted from NCBI backbone (NCBIN:108114, NCBI:P108115)

Query Match 4.1%; Score 208.5; DB 2; Length 2829;  
 Best Local Similarity 18.1%; Pred. No. 0.22;  
 Matches 215; Conservative 197; Mismatches 401; Indels 373; Gaps 53;  
 QY 23 EVFSLVKEPILKQTOASSISGADYAESGSKLKNINETSQVGVDTVDLPSDK-RTTPE 81  
 DB 188 QFNILILPKDSQYHT-----YEEET--LKQVTEYTNLSKTLMDSCISEKQDMIL 239  
 QY 82 KIKDNLAK-----GPREQLKAVTB-----NTESEKQITSGSQLESKESLSLNTV 128  
 DB 240 EYEINVAKRKSTETETLGDGKTELKSAVSQAYAKHLESYGVLPKPLNDIKNNAFSLVLDK 299  
 QY 129 PSTSNWEIC-DPITKNTLVGLSKSGVEKLSQTDHLVLPQAADGTQLIQVASFAFTDPK 187  
 DB 300 YCKDN---CGEYVQYNTMRKNFISSSDQYKMEAVVYIPKSIDNVTYLDKILS----- 349  
 QY 188 KTAIEYTSRAGENGEISQLDVGKEINE-----GEVNSYLLKKVITPTGYKHIG 239  
 DB 350 -----ESNELGIDIQETVNSLSKLGEBISEVSHLYVINSTLTDAAKKL- 393  
 QY 240 QDAFVNDKNIAEVLNLPESLETISDYAFALHALKQIDLPDLNKAIGELAFFQNGIKLSL 299

DB 394 -ESINEDESABEIDLQK-----PEDNSKALANNYCIFQYI----- 427  
 QY 300 PROLMRLAEAFKSNHNIKTIFRONSILKVGIEA--SFQDNDLSQMLPDGLEKIESEAPT 357  
 DB 428 -KTLNEPIKAYESKIVKSNELLSTIIDTLGKSATALQESFTDQ-----EECNKIKTEA-- 480  
 QY 358 GNPQDDHYNNRVVLTKSGKNPSGLATENTVYVNDKSLMQESPE----- 401  
 DB 481 -----EKVDDAEDICEKN-----EQIYEIPESEDETIDDKINDLQDL 519  
 QY 402 IDYTKWLEEDPTYQKNSVTGFSNKGLOKVKGNKHL-EIPKQHNGVTITEIG--DNAPFNV 458  
 DB 520 IDQMKYKDEIV---NNSEFISN-----RYKNIYENLXETYTELTENDIGKLENDTSKV 569  
 QY 459 DFQNTLRKVDLEVKVLPSTIRKIGAF-----AFOSNNLKSPFA 497  
 DB 570 NFYLQIRKINTEKTKIDESLQTVKEKFIKLSKEKIELKIEFKSVTEINRLQGES 629  
 QY 498 SDDL--BEIKEGAFMNNRIETLE-LKDKLVITIGDAAPHINHIYAIVLPESVQEI-GRSAF 553  
 DB 630 ARDLHEBQIKB--ILDKMAKKVHYLKELLKSLGKSSVYFTFEMNELLTASYDNNMEGFSAK 687  
 QY 554 RQNGANNLIFMGSKVKTILGEMAFI-----SNRLEHLDLSEOKQLTE-----IP 596  
 DB 688 KEKADNDINALYNSVYREDINALIEVEKFTVENKESTLEMLKDEEMEEKLQDAKETPAK 747  
 QY 597 VQAFSDNALKEVLLPASIKTIREBAFKKN-----HLKQLEVASALSIAFNALDDN 647  
 DB 748 LNFVSDDKLTDTVYKMSAEVNTNABGIKKEIAQKQFENVVHKMKKEFSDAFS-TKFEALQNS 806  
 QY 648 -----DGD-----EOPDNKVVK-----THNSYALADGEBF 674  
 DB 807 MQQYNQBGDAIEKHQNRSEKBEYFKNESVEEDLSRETEEBEYTHKHNNFSRRKGB-- 864  
 QY 675 IVDPKLSSTIVDLKILKLEGLDY-----STLR 704  
 DB 865 -----ISAETNNREVINKIESQLYNYGVIEKFSLIGDQNEVSTAKALKEKIVSDSLR 918  
 QY 705 QTT---QTQFRDMTTA-----GKALLSKSNLRQGEKQKFLQBAQFFLGRVDLJK 750  
 DB 919 DKIDQYETEFKEKTSVAVENTVSTIQSLSKAIDSLKRLNGS-----INNCKKYNTDIDLLR 973  
 QY 751 AIAKAEKALVTKATKNGQ-----ILERSINKAVLAYNNSAIKKNVVRLE-----KE 798  
 DB 974 SKITLREEVQKEMPKRKGDKCGENTALLLSLDRKMKKIN-----EKLNDRLNSLDTKK 1029  
 QY 799 LDLLTGLVEGKGLAQATMVQVYLLKTLPL-----PLPEYIYGLNVYFD----- 842  
 DB 1030 EDLLKFYSESKSIHLSDKQKGP--QDPLNRIDEMEDIKEDVDDELNVNYQVISENKVTL 1086  
 QY 843 -KSGKLIYALDMS---DTIGEG---QKDAYGNPI-----LNVEDNEGYHALAVATLAD 889  
 DB 1087 FKNSSVTYIEAMHSHINTVAHGITSNKNELKSVKEVEDKLNLFVEQNEEDYK--KYGNPEN 1144  
 QY 890 YEGLD-----TKTILNSKLSQTSIRQVPTAAYHRA-----G 921  
 DB 1145 EKQLEATRGMSKLUKEVINKHVSEMTQLESTANTLKNNAKENEHDLLELNKTKGQMRD 1204  
 QY 922 IFQAIQNAAAAEQLPKPGTHSEKSSSSSSANSKDRGLQSNPKTN 967  
 DB 1205 IYEKLIKIAEEL-----KEGTVNELKDANEKANK-----VEPEPERN 1241

## RESULT 12

T28677  
 rhoptyr protein - Plasmodium yoelii  
 C1Species: Plasmodium yoelii  
 C1Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C1Accession: T28677; C45521  
 R1Keen, J.; Sinha, K.; Brown, K.; Holder, A.  
 Mol. Biochem. Parasitol. 65, 171-177, 1994  
 A1Title: A gene coding for a high molecular mass rhoptyr protein of Plasmodium yoelii.  
 A1Reference number: 220508; MUID:95021522; PMID:7935623

A:Accession: T28677  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2269 <KE2>  
A:Cross-references: UNIPROT:Q26223; EMBL:L27838; MID:g457145; PID:g457146; PIDN:AAA21304  
R.Keep, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.  
Mol. Biochem. Parasitol. 42, 241-246, 1990  
A:Title: Identification of the gene for a plasmodium yoelii rhoptry protein. Multiple co  
A:Reference number: A45521; MUID:91101660; PMID:2270106  
A:Accession: C45521  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 2131-2269 <KE2>  
A:Cross-references: GB:M34283

Query Match 4.1%; Score 206.5; DB 2; Length 2269;  
Best Local Similarity 19.7%; Pred. No. 0.19;  
Matches 198; Conservative 169; Mismatches 369; Indels 267; Gaps 48;  
QY 40 SSISGADYAESGSKLKINETSGPVDVTDVDFSDKRTTPEKIKDNLAKGPREQELKAV 99  
DB 963 SNFNLEIMADNSQVILEIKKNGTNDHD---YNIKELSKHDKSGYKTEADQKKAI 1018  
QY 100 TENTSEKO-----ITSGSQLEQSK-ESLSINKTVPTSPSTNWEICDFITKGN 144  
DB 1019 QKNKELFEQYKEVTVLLNKYAVELKNFKDKNDKQIIEIKDAHNY--CT-LESKG 1075  
QY 145 TLVGLSKSGVEKLSQTDHLVLSQAADGTQLIQVASFAPDPKKTAAIAYTSRAGE--- 200  
DB 1076 SEKOMEIKKIHIEDEVANDNDKSKAITSIKVSVEPF-KTKIINKINEIRTKSDCLKE 1134  
QY 201 ----NGEISQLDVD-----GREIINEGEVFNLSLLKVTIPGKXHGIGODAFVDMK 247  
DB 1135 TNDLEKQINLSIDTQETKLTENGKQLTLELLESLLKQKKNI-----ED---QKK 1183  
QY 248 NTAENVLPSELTISDYAFALHALKQIDLPDLNKAIGELAFPDNQITGKLSLPLQMLRLA 307  
DB 1184 ELDEVN--SKINKIENTVNHQKNYKVEIGKINEIAKTN--KNQIESTKELIKPTIQHI 1239  
QY 308 ERAFKNHKTTEFRGNSKLVIGEAS--FQDNDLSQMLPDGLEKIESEAFNPGPDH 365  
DB 1240 ISSFNANDLEGDSDELNGKNTMGNIYEEPIKSYNLTNYLETYSKESITYN---QIQ 1296  
QY 366 NNRVLTWTKSGKNPSGLAENTVNPDKSLWQESPEIDYTKWLEEDFTQKNSVTGFSNK 425  
DB 1297 NKRIDTQKELLN-----IEN--VNAKSY-----LDYIK--ENEF--DRIVTHFKKK 1338  
QY 426 GLQVKRNKNLEIPKOHNGVTITEGDNAFRNVDFQNKTLR-----466  
DB 1339 -LNTVNDNPKNEYSKVGNEFDNISINTVKNSDENSELNLINQTKEMYANIVNNTYYS 1397  
QY 467 -KYDLEEV-----KLPTTRKIGAFQSNLKSFEASDDLEIKGAPMNNRIETLEUK 520  
DB 1398 KYEAEANIERNIPKLANLTN-----IKIKNSGIDLSKDI-KIATLSYLSKTE----- 1445  
QY 521 DKLVIT-----GDAFHNHIVAVLP--ESVQIGRSAPFRQNGANNLIF-----MGSKV 568  
DB 1446 DTLIFIPSPQKTTETTKISDSYSLDILKKSQELQK---KEQOQLKLIENRRLYEKV 1502  
QY 569 KTLGEMAFLSNRLHLDLSEQQLTEIPVOAFSDNALKE-----VLLPASIKTLR 618  
DB 1503 QATNE---LRGTSLDLKYKKEKILSEVKLLHKSNELNKLSCNFQNYDTILSSKYDQVK 1559  
QY 619 ERAFKQHLKQLEVASALSHIAFNALDNDGDEQFNKV-VVKTHNSYALADGEH---- 673  
DB 1560 E---KSNYKQEKELG---IDFNV---TDMEKFNNDIKVIEELENNYSDSENNNIQ 1610  
QY 674 -----FIVDPKLSSTIVDLEKIL-KLIEG-----LDYSTLTQTTQTQPDMT 715  
DB 1611 SKQKUKELTNKFAETKIKDDKIIIEKNDLIDKLIETKRNKMLFTHTTAEITLIKITDYS 1670  
QY 716 TAGKALLSKNSLRQGEKQKFLQEAQPFLLG-----RVDLDKAIKAKAEKALV 760

DB 1671 ---KFIESAT-----KFSKEFLKYIGDTSNLSNDDIATLQLKYDLHQ--INKYVTSKL 1718  
QY 761 TKKATKNGQLLERS-----INKAVLANNNSAIKKA--N 791  
DB 1719 SDATNNDNNLIEKEATEQAQKNTKLTFTIDSNNDNANALHNKTKQWYFNSLHKSIES 1778  
QY 792 VKRLKEKLDLTLGLVEGKPLAQATWQGVYL-----LKTPLPLPEYIYGLNV----- 839  
DB 1779 IKQLYKKMHVFKLLNIGQ-----INGKYPDISKQFDNLIQOESELTANLNDLKEIG 1830  
QY 840 ---YFDKSGKLIYALDMSDTIGEGQKADYCNPTILNVDNEGYH 880  
DB 1831 QKISDKKNKFLHALNET-----PIPNFTLKEIYH 1860  
RESULT 13  
D90011  
FntB protein [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: D90011  
R.Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A69756; MUID:21311952; PMID:11418146  
A:Accession: D90011  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2481 <KUR>  
A:Cross-references: UNIPROT:Q990K6; GB:BA000018; PID:gl3701961; PIDN:BAB43253.1; GSPDB:  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: fntB(mrp)  
Query Match 4.0%; Score 204.5; DB 2; Length 2481;  
Best Local Similarity 18.5%; Pred. No. 0.27;  
Matches 206; Conservative 159; Mismatches 403; Indels 343; Gaps 45;  
QY 10 LTLTVSVVTHNQEVSLVKEPIKQ-----TOASSSI-----SGDYAESGSKSLKI 58  
DB 1293 LQATQTVNNNDQANTBEKEAAIQALATAVTDAKNNITAAATDDNGVDATAKAGNSIQS 1352  
QY 59 NETSPVDDTVTDLSDKRTTPEKIKDNLAKGPREQELKAVTENTSEKQITSGSQLEQS 118  
DB 1353 TOPATAVKNNAKNEVDQAVTTQNAIDNTT-----GATTEKNAKDLVLKAKKAY 1404  
QY 119 KESLSNKT--VPSTSNWEICDF--ITKGNLVLGSKSGVEKLSQTDHLVLPSSQAADGTQ 174  
DB 1405 QDILNAQTNDVTQIKDQAVADIQGITADTIKQVAK-----DELATKAN----- 1449  
QY 175 LIQVASFAPDPKKTAAIAYTSRAGEISQLDVKGKEIINEGEVFNLSLLKVTIPTG 234  
DB 1450 -----EQKALIAQTADATTEKEQANQVDAQ-----LTQG 1480  
QY 235 YKHICQDAFVNKNIAEVLNPLESLETISDYAFALHALKQIDLPDLNKAIGELAFDNQIT 294  
DB 1481 NQNIENASIDDVNTAKDNAIQADFIQ-----ASTDVKTWARA-ELLTEMQNKIT 1530  
QY 295 GKLSLPLQMLRLAERAFKSNHKTTEFRGNSKLVIGEASQDNDLSQMLPDGLEKIESE 354  
DB 1531 EIL-----NNNETTNEEGNDIGPV-RAAYE-----EGLNNINAA 1564  
QY 355 APTGNPGDDHYNNRVVLTWTKGNPSGLATENTYVNPDKSLWQESPEIDYTKWLEBEDFTY 414  
DB 1565 TTGTD-----VTTAKDTAVQVQQLHANPVKKPAGK-----KELQQAAD 1604  
QY 415 QKNSVTGFSNKGLOKVKRNKNLEIPKOHNGVTITEIGNAFRNVDQFNKTLKYDLEVK 474  
DB 1605 KKTQIETQTPNASQOEINDAKQ-EVDTELN-----QAKTNVD-QSSTNEYVD----- 1648  
QY 475 LPSTIRKIGAFQSNLKSFEASDDLEIKGAPMNNRIETLE--LKDKLVITIGDAAPH 532

```
Db 1649 -----NAVEGRKAKINAVTFSEYKDKALAKIEDA--- 1678
Qy 533 INHIYAVLPBSVOEIGRSAPRQGNANLIPMGSKVTLTGEMAFVLSNLEHL-----DL 586
Db 1679 -----YNAKVNEADN-----SNASTSSBIAEAKQKLAELKQTADQNV 1715
Qy 587 SEQKQLTEIPVQPSD-NALKEVLLPASLK-----TIBEAFKK 624
Db 1716 NOATSKODIEVQIHNDLNDINDYIPTGKESATDLYAYADQKKNISADTNATQDEKQ 1775
Qy 625 NHLQLE--VASALSHIAFNALDNDGDE-----QFNKVVVVTKHNSYALA- 669
Db 1776 QAIKQVDQVQTALENIN-NGVDNGDVEDDALTOGKAADITQVDATVPKANQAIEAKAE 1834
Qy 670 DGEHFIVDPDKLSS--TIVDLEKILKLEGLDYSTLROTTQTQFRDMTTAGKALLSKSNL 727
Db 1835 DTKEISIDHSDQLTABEKTALAMIKQITDQAKQGITDATTAEVEKAKAQGLEAFDNIQI 1894
Qy 728 RQGEKQKQLOEAQFLGR-----VOLDKAIKAEKALVT-----KKATKNGQLLERSIN 776
Db 1895 DSTEKQKABELETALDQIRAGVNVVDATTEEKEAFTNALEDILSKATED--ISDQFTN 1952
Qy 777 KAVLAYNNSAIKKANVKRL-----EKELDLLTGLVEGKGPLAQAATMVQ 820
Db 1953 AEIATVKSALAEOLKAQRIINPVVKNALEATREVVNKQIEII-----KNADADASAK- 2005
Qy 821 VYLLKTPPLPEYYIYGLNVPDK-SGKL-----IYAL-----DMS 854
Db 2006 --IART-----DLGRYFDRPADKLDKTQNTNEVAELQNVITPAIEAIVQNDPDAN 2054
Qy 855 DT-IGEGOKDAYGNPILNVDENEGHALAVATLADYEGLDIKIILSKLSQLSIRQVP 913
Db 2055 DTNNGTDNDNATNSANATPENTQOPNVSETT--DNGKADASPTTPNNSDAATGETTVT 2112
Qy 914 TAAVHRAGIFOAIONAAEAELQLP-----KPGTHSEKSSSES 952
Db 2113 SATDDAKDKPQANNSSADASTNSTMDNDVTSRPEVESTNGTIDKPVETDNATPAES 2172
Qy 953 A---NSKDRGIGOSNPKTRNGRHSAILPRTGS 980
Db 2173 TTNNSTTTATNENAPTG---STATAPTAS 2200

RESULT 14
T31105
Hypothetical protein 2 - Haemophilus ducreyi
C:Species: Haemophilus ducreyi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T31105
R:Ward, C.K.; Lumley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
J. Bacteriol. 180; 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A:Reference number: Z20984; MUID:99030326; PMID:9811662
A:Accession: T31105
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4919 <WAR>
A:Cross-references: UNIPROT:Q9ZHL0; EMBL:AF057696; NID:g3929021; PID:g3929023; PIDN:AAC7
C:Genetics:
A:Gene: lepA2

Query Match 4.0%; Score 204.5; DB 2; Length 4919;
Best Local Similarity 20.0%; Pred. NO. 0.75;
Matches 250; Conservative 181; Mismatches 415; Indels 405; Gaps 65;

Qy 10 LTLTVSVVTHNQEVFSIVKPEILKQTOA-----SSISGADYABSSGSKLK----- 57
Db 598 VTLSKSTLSAGELTFKKVKVNTLNNDSELANNLSNASHVNTLNKSKUSAKADIKAK 657
Qy 58 -----INETSQPVDDTVTDLFSDKRTTPSKI-----KNL----- 87
Db 658 VNLTLNDTTELAKNLDINSITITNNGTIAGIFAN--ITTEKLNKKEKALILAEQNLNPT 715
```

```
Qy 88 -----AKPREQELKAVTENTSEKQITSGSOLEBQSKESLSLN----- 125
Db 716 VNGSHYENKGDIVSKDKATVTFPSKNSDFTSGSKLVNAQNQLKVNNVNFTISQGDITLI 775
Qy 126 -----KTVPSTSNWEICDPIFKGNTLVG-----LSKSGVE---KLSQT 160
Db 776 GNVTLNASGTFNTSGNLTITVTKLDVGDIQNTFNKGNLTVGEDLHKSKTKITNDGKLISI 835
Qy 161 DHLVLPQOA---ADGTOL-IQVASF---FTPDKKTAAIABYTSRAGENGEBISOLDV--- 209
Db 836 KNLNISSEADFINNGTLLGIEALKIATKGNFTNKEKAILASN-----SLLDISVA 885
Qy 210 DGRKEIINEGEVFNLYLKKVTIPTG-----YKHQDAPFVDNKNIAEVL----- 254
Db 886 EGKKTFFNG-----TIESGNLNTTNTGAFNLVDNATIRSFVLNITSTGNVSN 934
Qy 255 -----PESLETSIDYAPAHALAKQIDLPD-----NLKAIGE 285
Db 935 NGTLISNERLNITSAANFTNESNGTVMNSGLNIIAQGNITNKNLIASRQQLNLTAVD 994
Qy 286 LAPFDNOITGKLSLPROLMRLA-----BRAPKSNHIKTIBFRGNSLKVIGESAFODNLSQ 341
Db 995 NITNDSNISNKIAVLHSLGNISLNSKQVYNLGEI-----YAGNNISV--KAHQLKNDV-K 1047
Qy 342 LMLPDGLEKIESEAF-----TGNPGDDHYNNRVVLWTSGKNPSGLATENTYVNPDKSLW 396
Db 1048 LMGDITTTKTEGGQASYKLYQASNGG--HFGND-----GSSGYSEGDNLNKGKPADLDNKLIT 1101
Qy 397 -QESPEIDYTKWLEEDFTYOKNSVTGFS---NKGLOQVKRNKNLEIPKQHNGVITTEIGD 452
Db 1102 VQRTGKI-----YAGRDUTFNKSNAGKSEIINRGTIINVKNKLSYD-----SD 1144
Qy 453 NAFEN-----VDFQNKTLR- KYDLEEVKLPSTIRKIGA-----PAFOS--NNLKFEASDD 500
Db 1145 VSPENNMQSKQVDLYTKI PEAKSDIE-----LTPKTNGTHPVYLNFKSNNEKKYRNSEN 1199
Qy 501 LEEIKGAFMNNRIETLEBKDKLVTDGAAPHINHIYAVILPESVQSIGRSAPFRQNGANN 560
Db 1200 TKNFK-----SIGDL--INEALSDSAPEAIE-----AYTSGSSSN 1232
Qy 561 LIFMGSKVKTLGEM-----AFLSNRLEHL-----DLSEQKOLTEIPVQAFSD----- 602
Db 1233 YINPVSYLAALAGNANSSNPHYLTALKHILGNCWQDDLKQENI--KVLQKQWEDFPKOK 1291
Qy 603 NALKEV-LLP-----ASLKTIREBAFKKHNLKQLEVASALSASHIAPNA 643
Db 1292 GASKMLDLYPNTDKKAKIFAGIIRNGNDTISDVSEDFKKYKSKFQNGEAWKNDTGDS 1351
Qy 644 LDDNDGDEQFNDKVVVTHH-----NSYALADGEHFIV-----DPDKLSS 683
Db 1352 YDSTKASEKY--KKVENVDHKENIDEHLNIGKHEITVPGVSPENLNKNMNDHQDPKLGE 1409
Qy 684 TIVDLEKILKLEGLDYSLRQTQTQFRDMTTAGKAL--LSKSNLRQGEKQKFLQEAQF 741
Db 1410 --IDKSIISELQAQPVVTEKSAARDSPRVNDNDKEALDNLYRTRLSYINQNYL-GAKY 1466
Qy 742 FLGRVLDLKAIAKAEKALVTKKATKN---GQLLERSINKAVLAYNNSAIKK-----ANVK 793
Db 1467 PFNQLDTE-----DKLKGIRKIGDNYFBHQILTRLTEK--VADNHLTLKHLHDIALVK 1519
Qy 794 RL-----BKELDLTG---LVGKGKPLAQ-----ATMVQGVYLLKTPPLPEYIYL 837
Db 1520 KLIDSASIQAKDLMLKVEALTKEQKDNLKEDIWVYVKTVEVNGEVL-----VPQVYL-- 1572
Qy 838 NVYFDKSGKLIYALDMSDTTIGEGOKDAYGNPILNVDS-----NEGYHALAVATLADYE-GL 893
Db 1573 -----AKQIEEVEKORGVTGTQIRA---GLIDVKVDVDRNTG-----TIAGYAVGL 1616
Qy 894 DIKT-----ILNSKLSQLTISRQVPTAAHYRAGIFQAIQNAAAEABQLPKPGTH--- 943
Db 1617 EAKNKLKNTGDILSRLSKL-----VGKKGLESTGVTVVDGTATKVKARIKSEGHIYL 1671
```

Db 788 ILRDTSSRLQIEKETVYHREOQSOLL : : : : : LNSLEFIKTNLSEMEGRQLBQRLL 839

Qy 583 -----HLDLSEQ--OLTEIPVOAFSDNALKEV-----LLPASLKTIREEA 621  
Db 840 DDTVRELAQQRRHFQEEREFRESINEFKQAETAIKLDKEEKQADKWQALTSVRBEL 899

Qy 622 FKXNHKLQLEVASALSIAFNALDDNDGDGFONKKVVVKTHHSNYALADGEHFTVDPKL 681  
Db 900 AEKV-NVNELSKGLQEVLTPTLNDN-----HITAANKRAREFELKDQA 943

Qy 682 SSTIVDLEKITKLIL--EGLDYSTLRQTTOFPRDMTTAGKALLSКСNLROGЕKOFLQEA 739  
Db 944 TVWIESUTKELAKTRHEHQEQFYKMOSAESEIRLHELHGELYSK----QEEEIUKLRESS 999

Qy 740 QFFLGRRVDDLKAIAKAEKALVTWKATKNQGOLLERSINKAVLAYNNSAIKANVRKRLEKEL 799  
Db 1000 E-----AELKTRISDLAEAMLSNVTEOSTVNGS-----COLLSAQ 1036

Qy 800 DLTLGVLEGKGPLAAQMVGCVLYLLTPLPEYYIGLVANVFYFGSGKLIYALDMSDTIGE 859  
Db 1037 DLLKSILLE-----KLTEANCRTIRTLSSENTSLVESLNAAEV--- 1072

Qy 860 GQXDACNPILNVDEDNEGHYHALAVATLADYEGLDIKTILNSKLSOLTSIROPVTPAAHR 919  
Db 1073 ----KYANGMIQ-----HSADIOELTRYKAEFFKA--NDELNQLSGRESLOQAY- 1116

Qy 920 AGIFQAIONAAAEBOLLPKPPTHSEK----- 946

Db 1117 ---DELLRSNAEBAOKLLDKERESEKRVSDLHMNSLNHDQIEALASKVLASQSQNP 1172

Qy 947 -SSSESANSKDRGLQSNPXT--NRGRHSAILPRTGSKGSFVYGILGYTSVALLSLITAI 1003  
Db 1173 NSSLINESAMDGQDSLASGLTAABEGRNNE-----QLLKIHKFL 1211

Qy 1004 KKKK 1007 :||  
Db 1212 RKEX 1215 :|:

RESULT 16

B89921

Hypothetical protein ebha [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C;Accession: B89921

R.;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui,  
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaiho, C.; Sekim  
C.; Shiba, T.; Hatatori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.

A;Reference number: AB9758; UID:21311952; PMID:11418146

A;Accession: B89921

A>Status: preliminary

A:Molecule type: DNA

A;Residues: 1-6713 <KUR>

A;Cross-references: UNIPROT:Q9U54; GB:BA000018; PID:g13701232; PIDN:BAB42527.1

A;Experimental source: strain N315

C;Genetics:  
A;Gene: ebha

Query Match            4.0%; Score 203; DB 2; Length 6713;  
Best Local Similarity    19.5%; Pred. No. 1.4;  
Matches     232; Conservative   183; Mismatches   396; Indels   376; Gaps   64;

Qy .34 KOTOASSISSGAD-----YAESSGGSKLKLNETSGP-VDDTVTDLFPSDKRTTPEKIDN 86

Db 5050 QQTEAGSKFINEDKPOKDAQAQVAQNAKOLINOTGNPTLDKSQVEQLTQAVTT---AKDN 5106

Qy 87 L---AKGPREQELKATEN-----TESEKQITSQSGLQESKE 120

Db 5107 LHGDQKLARDQQCAVTVTVALPNLHAQQOTLTDAINAAPRTREVAQHVOGTATELDHAME 5166





```
Db 1087 LDANSVDDDKAAIKESDAMQF---DSVQDLNGDIFNN-----TKDQVAPLND----- 1132
Qy 534 NHIYATVLPESVQIEGRSFRONGANNLIIFMGSKVTILGEMAFSLNRLEHLDLSEKOLT 593
Db 1133 -----LLEKAE-GRS-ISAENATLI---QKDELAQAISIEGVVVKINRDEVIKOR 1180
Qy 594 EIPVQAFSD-----NALK-----EVLPLASLKTREE----- 620
Db 1181 KVKLDAYNDMVYTSNKLMTKEVNNNAIKTLNADTLRIDSLKLKRLKLDWSBAELSDLEV 1240
Qy 621 -----AFKKNHLKOLE-----VASAL-SHIAFNALDNDGDEQPD 654
Db 1241 KSINNVDADAKELKLEEKMLQPGGYSNQSIEAMQSVKSALESYI--SASEEATSTQEMN 1298
Qy 655 NKVVVTHNSVALADGEHFIVDPDKLSSTIVD-----LEKILKLEGL-----DYSTLR 704
Db 1299 KQALVEAGTSLNWTDOQKANEETKTSYVVDVKYKALEKVNAEIDKYNKQVNDYPKYS 1358
Qy 705 OTTOTQFRD-MTTAGKALLSKNLRQGEKQKELQEAQFFLGRVLDLDAKAEKALVTKK 763
Db 1359 Q-----KYRDAIKKEIKALQKKLMQ-EQAKLLKO-QIKSGNI-----TQYGIIVTST 1404
Qy 764 ATWNG--QLLERSINKAVLAYNNSAIKKANVRLEKELDLTLGLVEGK-GPLAQAATWVG 820
Db 1405 TSSGGTSPSTGGSYSGYSKYNSAASKYNV-----DPALIAAVIQOESGFNAKARSVG 1459
Qy 821 VYLLKTPLPEYIYGLNVVFPKSGKLIYALDMSDTIGSGOK-DAYGNPILNVDENEG 879
Db 1460 AMGLQMLPATKSLGNNAYDP-----YQNVGGTKYLAQOLEKPGG--NVEKALAA 1511
Qy 880 HALAVATLADYEGLD-----IKTIL---NSKLSQLTISRQVPTAAVHRAGIFQAIQN 928
Db 1512 NA-GPGNVIKYGIIPFKETQNVVKKIMANYSKLSATS-----SIASY----- 1556
Qy 929 AABAEQLLPKGTISEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG--SFVY 986
Db 1557 -----TNSAPRVSSKYQOESGLRSSP--HGKTDFAAKAGTAIKSLQSGKV 1601
Qy 987 GILGYTSA 995
Db 1602 QIAGYSKTA 1610

RESULT 18
H89960
hypothetical protein SA1577 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: H89960
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A:Reference number: A89758; PMID:21311952; PMID:11418146
A:Accession: H89960
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2186 <KUR>
A:Cross-references: UNIPROT:Q99TB0; GB:BA000018; PID:g13701552; PIDN:BA842845.1; GSPDB:G
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA1577

Query Match 3.9%; Score 199; DB 2; Length 2186;
Best Local Similarity 19.2%; Pred. No. 0.4;
Matches 217; Conservative 182; Mismatches 387; Indels 342; Gaps 58;
Qy 20 HNOVFSLVKEPILKQTAQSSISGADYABSSGSKLKNETSGPVDDTVTLDFSD---- 75
Db 1208 YNAKLAENATPDATNDEKNAINTLNQDROQAIESIKQANTNAEVDQAATVAENNIDAV 1267
```

```
RESULT 19
P84730
probable myosin heavy chain [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84730
R:Lin, X.; Kaul, S.; Rounalev, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.
Nature 402, 761-769, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: F84730
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1269 <STO>
A:Cross-references: GB:AE002093; NID:G6598483; PIDN:AAC69932.2; GSPDB:GN00139
C:Genetics:
A:Gene: At2G32240
A:Map position: 2

Query Match      3.9%; Score 197.5; DB 2; Length 1269;
Best Local Similarity 20.3%; Pred. No. 0.21;
Matches 235; Conservative 167; Mismatches 440; Indels 313; Gaps 51;

Qy      1 MKKHLKVALTLTTVVTHNQEVSLVKEPIKILKQTOASS-----ISCADYAE 49
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
270 MSENEKVEAALKSSAGALAAQVEALSKSLRLTEQKVSSTEALIDELTQEOEKKASE 329
Qy      50 SSGKSKLINETSGPDDTVTDLFSDKRTTPKIKDNIAKGPQEQELKAVTENTSEKQI 109
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
330 SPKKELSVLQDL----DAQTGLQAKLSEQEGINSKLABELKEKELLESLSKQDEKLR 385
Qy      110 TSGSQLEQ----SKESLSLNKTVPTSN----WEICDPTFKNTLVGLSKSGVEKLSQTDH 162
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
386 TANEXLAELVLEKEALEAN-VAEVTNSVATVTEVC-----NELSEKLTSDENFSKTD 438
Qy      163 LVLPSQADGTQLIQVAFPTPKKTAIAYTSAG-----ENGEI-----SQL 207
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
439 LL--SQALSNNSELE-----QKLKSLLELHSEAGSAAAAATQKLELEDVVRSSQA 488
Qy      208 DVDGKEIINE-----GEVFNLYLKVTIPTGVKHIGQDAFVNKNIAEVLNPLESLET 260
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
489 ABEAKSQIKELETKTAAEQKVAEEQQLNL---LQLKSSDAERELKELSEKS--SELQT 543
Qy      261 ISDYAPAHALAKQIDLPNLKAIGBLAFPDNQITQKLSPLQMLRLA-----BRAPKS 313
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
544 AIEVAEEKKQATTQMEYKQKASELSLSLTQSSARNSELEEDLRIALQKGAHEDRANT 603
Qy      314 NHKITEIFRG---NSLKVIGEASFDNDLSQLMLPD-----GLEKIESEAFGTG 358
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
604 THRSIELEGLCOSSQKHEDAEGLKQLELLQTEKYRIQELBQVSLSEKKHGETRAD 663
Qy      359 NPGDDHYNRV-----VLWTKGKNPSG--ATENTYVNPDKSLMQESPEIDYTKWL 408
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
664 SKG---YLQVAVELOSTLEAFQVKSSSLEAALNIATEN-----698
Qy      409 EEDFTYQKNSVTGFNGKGLQKVRKNKLEIPKOHNGVITTEIGDNAFRNVDPQNKTLRY 468
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699 EKELTENLNAV-----SEKKCLEATVDSEVSKISE--SENLESLE-----RNEL 741
Qy      469 DLEEVKLSTIRKIGAFAPQSNILKS--FEASDDLEEIKEGAPMNRRIETLELKQKLVTI 526
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
742 NVTQKKLES-----ENDLKAAGLQSEVMEKLSAE-----ESLEQKGR--BI 783
Qy      527 GDAAFPHINHIVAI-----VLPEVQVEI--GRSAFRQNGANNLIIFWGSVKV-----569
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
784 DEATTKRMELEALHOSLSDSHRLQKAMEEFTSDSEASSLTKELRDLEGIKSYEQRL 843
Qy      570 -----TLGMAFLSNRLHLEL-----DLSEKQL-----TEIPVQAFSDNALK 606
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
844 ABASGKSSSLKEBQTLGLRLAAESVNEKLAQEFDAQEKSLQSSSESELLAETNNQLK 903
```

```
Qy      607 -----EVLTPASLKTIREEAFFKNHKLQLEVASALSHIAFNALDDNDGDGFQFNKVVVK 660
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
904 IKIOELSLGSG-----SVEKETALKRUEEAIE----RPNQKE-----TESSDLVEKJK 949
Qy      661 THNSY----ALADGEHFIVDPDKLSSTIVDLKILKILIEGLDYSTLR-----QTTQTQ 710
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
950 THENQIEYKKLAHEASGVADTRK-----VELEDALSKLNLE-STIEELGAKCQGLEKE 1003
Qy      711 FRDWTAGKALLSKSNLRQGEKQKFLQBAQFFLQGVLDLKAIA-----KAEKALV---TKK 763
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1004 SGDLAEVN-----LKNLELANHGSEANELOTKLSALEAEKEQTANELEASKTTIEDTKQ 1059
Qy      764 ATKNGQLERSINKAVLAYNNSAIKKANVKELEKEL-DLLTGLVEGKPLAQATMVQGVY 822
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1060 LTSEGEKLSQIEKLRVAEAEKSVLESHFELEKTLSEVKAQLKENVENATASV-----1114
Qy      823 LLKTPFLPEYIIGLVYFDKSGKLIYALDMSDITIGGQKDAYGNPIINVDNEDNEGYHAL 882
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1115 -----KVAELTSKJQEHIEHIA-GERDVLNEQVLQKELQA-----1149
Qy      883 AVATLADYEGLDIKTLNLSKLSQTSIRQVPTAAHYHRAGIPQAIONAAAABEQLLPKPGT 942
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1150 AQSSIDSQKRLKRLKSLNLNQ-----WSKILNRKCSSQML-----KL 1188
Qy      943 HSEKSSSESANSXDRGIQSNPKTR--GRHSAILPRTGSKG-----SFVYGI 988
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1189 RLETEADVGVKSRIIDLSFSPPTKRKSKKPEASLSSSSSGNVTTPTQTASTSHLMTV 1248
Qy      989 LGYTSVALLSLITAI 1003
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1249 KIVTGVALLSVIIGI 1263

RESULT 20
S67593
transport protein USO1 - Yeast (Saccharomyces cerevisiae)
N:Alternate names: protein D2552; protein YDL058w
C:Species: Saccharomyces cerevisiae
C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004
C:Accession: S67593; A38455; S30782
R:Bloecker, H.; Brandt, P.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67587
A:Accession: S67593
A:Molecule type: DNA
A:Residues: 1-1790 <BLO>
A:Cross-references: UNIPROT:Q07380; EMBL:Z74106; NID:G1431058; PID:G1431059
A:Experimental source: strain S288C
R:Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamaeaki, M.
J. Cell Biol. 113, 245-260, 1991
A:Title: A cytoskeleton-related gene, USO1, is required for intracellular protein transp
A:Reference number: A38455; MUID:91185402; PMID:2010462
A:Accession: A38455
A:Molecule type: DNA
A:Residues: 1-389, 'TA', 392-724, 'S', 726-1790 <NAK>
A:Cross-references: GB:X54378; NID:G4777; PIDN:CAA38253.1; PID:G4778
A:Note: the authors translated the codon ACT for residue 768 as Ile
R:Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.
submitted to the EMBL Data Library, February 1993
A:Description: An integrin analogue in Saccharomyces cerevisiae.
A:Reference number: S30782
A:Accession: S30782
A:Molecule type: DNA
A:Residues: 71-846, 'E', 848-923, 'K', 925-1252, 'I', 1254-1318, 'V', 1320-1460, 'S', 1462-1580, 'S'
A:Cross-references: EMBL:L03188
C:Genetics:
A:Gene: SGD.USO1; INT1
A:Cross-references: SGD:S0002216; MIPS:YDL058w
A:Map position: 4L
C:Keywords: coiled coil; transmembrane protein
P:326-342/Domain: transmembrane #status predicted <TM1>
P:394-410/Domain: transmembrane #status predicted <TM2>
P:617-633/Domain: transmembrane #status predicted <TM3>
```

Query Match 3.9%; Score 197.5; DB 2; Length 1790;  
Best Local Similarity 21.0%; Pred. No. 0.35;  
Matches 248; Conservative 174; Mismatches 425; Indels 333; Gaps 59;

QY 25 FSLVPEPILKQTOA---SSSISGADYABSSGSKLKNITSGPVDVTTD-----71  
DB 434 FLQORDFLKQVLCNNSTNNVGDNAKENGSKNSDKESDQDGTGDTGTEGYSFKAN 493  
QY 72 LF-----SDKRTTPEKI-----KDNLAQPREQELKAVTENTSEKQITSG 112  
DB 494 LFEVLLNYDALNLPKLFPTTDFPMFFQDQH-----KYSELREITRN-----VTTG 543  
QY 113 SOLEQSKESLSLNTVP---STS-----NWEICDPITKGNLTGLSKS 152  
DB 544 NDLE-DEEPLKAIQITISELTTSLTAADIRIPISYLTFLIYWLPGDFKATNDFL--SDKS 600  
QY 153 GVBKLSQTDHLVPSQAD-----GTQLIQA---SFAFTDDKTAIAEY--TSRAGEN- 201  
DB 601 VIKSLLSFSYQI---QDEVDTIKCLVTMLGVAYEFSSKSPFPRKEYFEFTIKTLGKDN 657  
QY 202 --DEISQDQD---GKEIINEGEV-----ENSYLLKKVT-----IPTGVKH 237  
DB 658 YASRIKQPKDGYFQKVDNNEISILTPDELGLPKVYFSTYFIQLFNENIYRIITLASH 717  
QY 238 IQQDAFVDNKNIAEVLNPE-----SLETISDYAFAPHLAKQIDLPDNLKAIGEL 286  
DB 718 DPDEBPINKISPEEVEKLQROCTKLKGEITSLOTETESTHENLTKLIAIWNHEKLEDEK 777  
QY 287 AFQDN-----QITKLSLQSLRQMLRAEPKNSHIKTIIFRGNLSKV 328  
DB 778 YOILANSSHSLKFNFSILETELKNVRDSDLEMTQLRDVLETKDQENQALLEYKSTIHK- 836  
QY 329 IGHASQDNDLSQMLPDGLKLESEAFNGCDHYNNRVLTWTSKGNQPSGLATENTY 388  
DB 837 -----QEDSIK--TLKGLTILSQK---KKAEDGIN-----KMGKDLFALSREMQA 878  
QY 389 VNPD--KSLMQESPEIDY-----TKMLEEDFTYQKNSVTGFSNKGKQVK--RNKNLEI PKQ 441  
DB 879 VEENCKNLQKEDKQNNVNHQETKSLKEDIKAIKAI--NENLEEMKIQCNNLSKEKE 937  
QY 442 HNGVITITEGD-----NAFRNVDFONTLRYKLDLEVKLPSTIR- 480  
DB 938 HYSKELVEYKSRFQSHDNLVAKLTKELKSLANNYKDMQAEENSLIK-AVEESKNESIQ 996  
QY 481 -----KIGAFQSNLKK-----SFEAS-----DDLEEIKEGAFMNNRIETLE- 518  
DB 997 SNLQNKIDMSQKENFQIERSIEKNIIEQLKTTISDLEQTKBEIISKSDSKDEYESQI 1056  
QY 519 --LQDKLVITIGA-AFPHINHIYAIVLPSVQIGRSAPRQNGANNL-IPMGSKVKTGLGM 574  
DB 1057 SLKKEKLETATTANDENVNKNISELTKTRELEALEAAYK-NLKNLETKLETSEKALKEV 1115  
QY 575 AFISNRLEHLDSEQKQLTEIPVQPSDNALKEVL-----LPASLTIREAFKKNHLK 628  
DB 1116 KENEELHKEEIKLEKEATEKQOLNSLRANLESKEHEDLAAQLKYYEEQIANKERQY 1175  
QY 629 QLEVASALSHIAFNALDNDGDEQDPNVKVVKTHNSYALADGEHFIV-----676  
DB 1176 NEEISQ-----LNDEBITQOENESIKKNDQ-----LEGEVKAAMKSTSEOSNLKK 1222  
QY 677 -DPDKLSSTIVLEK-----ILKLI EBLGYSTL-----ROTTOTQFRDMMT 716  
DB 1223 SEIDALNLOIKELKKNETNEASLIESIKVSESETVKIKELQDECNPFKEVSELEDK 1282  
QY 717 AGALLSKSNLROGEKQKFLQ-----BAQFPLGRV-DLDRKAIKAEKALVTKKAT--- 765  
DB 1283 ASQDKNSKYLEQKSEKIKELDAKTTELKIQLEKIITNLSKAKESESELRLKTSSE 1342  
QY 766 --KNG-QLLERSINKAVLAYNSAIKANVKLEKELDLITGLVEGKGLAQATVQGVY 822  
DB 1343 ERKNAEQLKELKNE--IQIKNOA-----FEKERKLLN---EGSSITIQ----- 1381

QY 823 LLKTPLPPEYVIGLVNVDKSGKL-----IYALDMSDTIGEGQKDAYGNPILNVDNEDE 877  
DB 1382 -----EYSEKINTLEDELIRIQENELKAKEIDNTRSELEKSVLSNDELSEKQN- 1431  
QY 878 GYHALAVATLADYEGDLIKTILNSKLSQITSI-----ROVPTAAYHRAGIFQAIQNAAE 932  
DB 1432 -----TIKSLOD--EILSYKDKITRDEKLLSIERNKRDLSLKEQ-----LRAAQESKAK 1481  
QY 933 AEQLLPKPGCTHSEKSSSESA---NSKD--RGLQSNPKTN 967  
DB 1482 VEEGLK-----LEESSEKKALEKSKEMMKKLESTIESN 1517

RESULT 21  
B70126  
surface-located membrane protein 1 (lmp1) homolog - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 16-Aug-2004  
C:Accession: B70126  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horset, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: B70126  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1119 <LE>  
A:Cross-references: UNIPROT:O51228; GB:AE001131; GB:AE000783; NID:g2688098; PIDN:AAC6659  
A:Experimental source: strain B31  
C:Superfamily: tetratricopeptide repeat homolog  
F:742-774/Domain: tetratricopeptide repeat homolog  
F:775-808/Domain: tetratricopeptide repeat homolog <TT2>  
F:809-842/Domain: tetratricopeptide repeat homolog <TT3>  
F:843-876/Domain: tetratricopeptide repeat homolog <TT4>  
F:911-943/Domain: tetratricopeptide repeat homolog <TT5>  
F:944-977/Domain: tetratricopeptide repeat homolog <TT6>  
F:979-1010/Domain: tetratricopeptide repeat homolog <TT7>  
F:1011-1044/Domain: tetratricopeptide repeat homolog <TT8>

Query Match 3.9%; Score 197; DB 2; Length 1119;  
Best Local Similarity 19.6%; Pred. No. 0.18;  
Matches 212; Conservative 164; Mismatches 358; Indels 347; Gaps 54;

QY 18 VTHQEVSLVKEPILKQTOASSISGADYABSSGSKLKNITSGPVDVTTDIFSDKR 77  
DB 188 VQDKNLFNLEK---LKK-----NLSGKSNENILNDSQKIE-----NDQK 225  
QY 78 TT---PEKIKDLAKGPREQ-----LKAVTENTSEKQITSGQLEQSK---ES 121  
DB 226 NTNLSKESNENILKTPDNSKYNNNTSLKSSNSQKSELSPPSQTIIGKIYRPVS 285  
QY 122 LSLAKTVPSTNWEICDPITKGNLTIVG-----LSKSGV-EKLSQTDHLVPSQAADGTQ 174  
DB 286 YLIKEL-----YEILDDINTGRVTLGKNLKLKLGKSNKFKQVKNELIENSKNEASN 340  
QY 175 LI-----QVASFAPTPDKKTAIAEYTSRAGENGESIQDVGKKEINEGEVFNYSY 224  
DB 341 LLLTLIKKIDIBPILNINIPKPYK-----EIFQLDKEDKK-----PQY 378  
QY 225 L-----LKKVTIPTGYKHIGQAFVD-----NKNIAEVLNLPESLETI 261  
DB 379 LEDLKSQVHSIKPIDLENTKSR--QQAIKDLNEFLKNNPNDAQASKTLAQAQKIQHLDL 436  
QY 262 SDYAFAPHLAKQIDLPDNLK---AIGELAFF-----DNQITGKLSLPRQLMRLAERAF 311  
DB 437 KSKVH---SIRPIDL-ENTKSRQQAIDKLNFLKNNPNDAQASKTLAQAQKIQHLEDLS 492  
QY 312 KSNHIKTIET-----RGNLSKVI GEASQDND-----LSQMLPDGLEKIESEAPT 357  
DB 493 KVHSIKPIDLENTKSRQQAIDKLN--FLKNNPNDAQASKTLAQAQKIQHLEDLSKVHS 550



Query Match	3.9%; Score 197; DB 2; Length 2441;
Best Local Similarity	20.8%; Pred. No. 0.59;
Matches	Conservative 141; Mismatches 375; Indels 270; Gaps 52;
Qy	70 TDLPSDKRTT-----PEKIKONLAKG-----PREQELKAVTENTESE-----KQITS- 111
Db	729 TDLAKGKEVTNKPHEVDEYDQSLAKGKOITNKPHE-----SVDEYDQITLAKGKEVTNK 784
Qy	112 -GSOLEQSKES-LSLNKTVPSTSNWEICDF-----ITKGNITLVGLSKSGVKLSQDTHLVL 165
Db	785 PHENLEEYNETDLAKGKEVTNKAHENLEEYNETDLAKGKEVTNKAHENLEEYNETD----- 840
Qy	166 PSQAADGTQLIQVASFAFTPKTKAIA-----EYTSRAGEN-GEISQLDV-DGKEIINEG-- 218
Db	841 ---LAGKEVTNKAHENLEEYNETDLAKGKEVTNKAHENLEEYNETDLAKGKEVTNKARE 897
Qy	219 --EVPNSYLLKVTIPTYGYKHIGODAFVDKNTAEVNLPESETISDYA-----FAHL 269
Db	898 NLBEYNETDLAK-----GKE-VTNK--AHENLEEYNET--DLAKGKEVTNKAHE 941
Qy	270 ALKQIDLPDNLKAIGELAFPDNOITGKLSLPQLMRLAERAFKSNHKTTFPGNSI--- 326
Db	942 NLBEYNETDLAK-----KEVTNKA-----HENLEEYNETDLAKG 976
Qy	327 -KVIGEASFQDNDLSQMLPDGLEKIESEAFGTGNPGDDHYNNRVLWTKSGKQPSGLATE 385
Db	977 KEVTNKAHENLEEYNETDLAKGKE-VTNKA---RENLEEYNETDL---AKGKEVTNKARE 1029
Qy	386 NTVYNPKSILWQSPEDIDYTKULE-----EDFTYQKNSVTGFSNKGLOKVKVRN 433
Db	1030 NL-----EYNETDLAKGKEVTNKAHENLEEYBEKDYMKON--ELQNKGSQGLKEN 1078
Qy	434 KNLKLEIPKQHNQVITTEIGDNAFRNVDFQNTLARKYDLEEVKLPSTIRKICAFQFSNNLK 493
Db	1079 AELKNKELRN-----KGSQGLKENAELKNKELRNKSGDGLK-----ENAEK 1120
Qy	494 SFEASD-DLBEIKEGAPMNNRIETLELQKLVITGDAAFHINHIYAVLPESVQEIGRSA 552
Db	1121 NKELQNKSGEGLKENAELKNK-----ELQNK-----GSEGLKEN---AELKNKELRNKSGEG 1169
Qy	553 FRQNG-ANNLI FMGSKVKTIGEMAFLSNR-LEHU-----DLSEQQLTETIPVOAFSDNALKE 607
Db	1170 LKNAELKNKELQNKSGEGLKENAELKNKELQNKSGEGLKENAELKNKELQNKSGEGLKE 1229
Qy	608 VLVPASLKT-----ITREEAFKNHLKOLEVASALSHPAFNALDDND-GDEQFDNK 656
Db	1230 ---NAELKNKELRNKSGEGLKENYVTNNDLKN-----NDIQNKDLNKNMKKN 1274
Qy	657 VVKTHNSYALADGEHFIVDPDKLSSTIVDLEKILIEGLDYSTLURQTQTQFRDMTT 716
Db	1275 ELLNKDLSNKMKNKE--LLNKOLSNEDMKKNKELLNKDINRKOLKSIGNWEQ-----QNT 1327
Qy	717 AGKALLSKSNLRQGEKOKFLQEAQFPLGRVLDOKATAKAEKALVTKKATKNGQLLERSIN 776



Db 372 DEDNEVLAKSSDFIFLKKQLIKERRTKKHEHLQNOIETFIIVELEHKVPIIINSFKERTDML 431  
QY 217 EGEVFNYSLLKKVTPTGYKHIGQDAFVDNKNIAEVNLPESLETISDYAFALHAKQIDL 276  
Db 432 ENELNNAALL-----LEHTSNE--KNAKVKELNAKQ-----KLVEB 466  
QY 277 PDNLKAIGELAFFNOITGKLSPLQMLRLAERAFKSNHIK-----TIEP----- 321  
Db 467 ENDLQTLTK-----QRDLQCRQIQIYLITNSVSDSGPLRKEBIEQIQINIMQSD 517  
QY 322 ----RGNSLKVIGASFOONDSLQW-----LPGLE-----KISEE 354  
Db 518 STITESDQKVTELVFKNIIQKQNAELLKVVRLNADKLESKEKSKQSLOKISE 577  
QY 355 AFTGNPGDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLMQESPEIDYTKWLEEDTY 414  
Db 578 --TVNEAKE---AIIITLSEK-----MDLESRIEELQKELEELKTS 613  
QY 415 QKNSVTGFSNGLQKVRNK-NLEIPKQHNGVTITEIGDNAPRVNDFONKILRYKDYLEV 473  
Db 614 VPNEADASYNVTIKQLTETKRDLESQVQDLOTRIQTIRESTENMSSLNKEIQ--DLYDS 671  
QY 474 KLPSITIRKIG-----AFAPQSNLKSFRASDDLEBIKEGAFNNRI----- 514  
Db 672 KSDISI-KLGKESSRILAEERFKLLSNTLDLTAKBND-QLRKRPDYQLNTILKQDSKTH 729  
QY 515 ----ETLEKOKLVITGDAAPHI--NHIYAIVLPEV--QETGRSAFRQNGANNLIPMSGK 567  
Db 730 ETLNEVSVCKSLIVETELLNKEEQKLVHLEKNLQELNKLSPK---DSLRIWYQ 786  
QY 568 VKTIGEMAFNLRLHLDLSEQKQTEIPVQAFSONALKEVLLPASLTIRREAFKQHL 627  
Db 787 LQTLQKER--EDLLETBTRKSCQKIDEL-----EDALSE-----LK--KETSQKH 830  
QY 628 KOLEVASALSHTAFNALDNDGDEOFDNKV--VVKTHNSYALADGEHFIVDPKLSSTI 685  
Db 831 KOLEE-----DNNSNIEWQNKIEALKKDYESTVSDSKQ--TDIEKQYKV 876  
QY 686 VDLEK-----ILKLEIG-----LDYSTLRQTT--- 707  
Db 877 KSLEKEIEBDKIRLHTYVMDETINDDSLRLKELEKSKINLTDAVSQIKEYKDYETTQS 936  
QY 708 --QT-----QPRDMTTAGKALL--SKSNLRQEGKQKFLQBAQFFL-GRVLDLKAIAKAB 757  
Db 937 LQOTNSKLDESFKOFTNQIKNLTDEKTSLE--DKISLLKQBFNPLNNELDLQKGMEXEK 994  
QY 758 ALVTKKATKNGOLLERSINKAVLAYNNSAIK-----KANVKRLEKELDLTLGLVEGKGP 811  
Db 995 ADFKURI-----SILQNNKEVEAVKSEYKSKISQINDLDOQT----- 1033  
QY 812 LAQATMVQGVYLLKTPPLPEYIIGLVYFVKSGKLIYALDMSDTIGB--GQKDAYGNPI 869  
Db 1034 -IYANTAQNY-----EQELQKHADVSKTISELREQLHTYKGV 1071  
QY 870 --LAVDED-----NEGYHALAVATLADYEGLDIKILNSKLSQLSIRQVPTAATHR 919  
Db 1072 KTLNLSRQLENALKENKSSSQKESL--LEQLDLS--NSRIEDLSSQNKL--LYDQ 1123  
QY 920 AGIFQAIQNAAAEQQLPKPQTHS-----EKSSSES 952  
Db 1124 IQIYTA---ADKEVNNSTNGPOLNILITLRRERDILDTKVTVAERDAKMLRQKISMDV 1180  
QY 953 ANSKDRGLQSNPKTNWRGHSAIL 975  
Db 1181 ELQDARTKLDNSRVEKENHSSII 1203

## RESULT 25

T18296

myosin heavy chain - Entamoeba histolytica

C:Species: Entamoeba histolytica

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T18296  
R:Guillen, N.  
submitted to the EMBL Data Library, February 1997  
A:Reference number: Z18865  
A:Accession: T18296  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2139 <GUI>  
A:Cross-references: UNIPROT:Q07569; EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB480  
C:Genetics:  
A:Gene: mhca  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 3.9%; Score 196.5; DB 2; Length 2139;  
Best Local Similarity 21.7%; Pred. No. 0.51;  
Matches 228; Conservative 121; Mismatches 369; Indels 331; Gaps 52;  
QY 1 MKXHLKTVALTITVTVSVVTHNQVPSLVKPEILKQTAQSSISGADYAESGSK---LK 57  
Db 1237 IEKEMKALQEBKENVESKNSKTEKDKKLELDNLDKDTQKLDMDTADNEKAKAKADLEAQ 1296  
QY 58 INETSGPVDDTVTD--LPSDKRTTPEKIKONLAKGPREQLKAVT-----ENTES 105  
Db 1297 LNEVDNHEKAVADAELLNKKKAQSDKELNSL-----KAELEALTAKSVVESKNKUSEN 1351  
QY 106 EKQITSSQLEQSKESLSLNTKVPSTNWEICDPITKGNITLVGLSKSGVEKLSQTDHLVL 165  
Db 1352 EKAALS-EEIDQANEKL---KNIQADLRKATAD-LQEBANE---KKAEBE--AQRDKLVA 1400  
QY 166 PS-----QAADGTQLIQVASPAFTPPDKKTAIAEYTSRAGENGESOLDVDGKEII 215  
Db 1401 DNKKQMTLBEIKARDEENTYKVENYKVLKRGKADLE---EAVEN-----LDIEKKDRM 1452  
QY 216 N-----EGEV-----FNSYLLKKVTPTGVKHIGQDAFVDNKNIAEVNLPESLETI 261  
Db 1453 NKEQVKKLEGELKETKDKLNAAIAEKOSIFTAKKQSDADLELNKTVEE----- 1502  
QY 262 SDYAPAHALAKQIDLDPNLKAIGBELAFPDNQITCKLSPLQMLRLAERAFKSNHIKTIEP 321  
Db 1503 HDEVVAKLNTQITKLTTR-----DNQ-----SABEELNELRSKADKDK----- 1539  
QY 322 RGNSLKVTIGRASFOONDSLQMLPDGLEKTESAFTGNPGDDHNNRVVLWTKSGKPSG 381  
Db 1540 ----KKISELEEQVNELES--RPVG-----TGN----- 1561  
QY 382 LATENTYVNPDKSLMQESPEIDYTKWLEEDFTYQKNSVTGPNKGLQKVRKNKLEIPKQ 441  
Db 1562 -ADENEIKIRDAQI-----ADLNKALE-----MKGVQNNQLQAT---NKLKAK 1601  
QY 442 HNGVTI--TEIGDNAPRVNDFONKTL-----RKVDLEEVKLPSTIRKIG 483  
Db 1602 DNDLTSKIEITENEMKKLENKAKRLEQDKDBADKAVSEQTIKRGLEBEVKKLT--BIQ 1659  
QY 484 AFAPQSNLKSFEASD-----DLEIKEGAFNNRIETLEKDKLVITIGDAAPHINHI 536  
Db 1660 ALKFOINAPSVQAESEKQRLSEDIABKE-----QLEOERTTAANAERAKKI 1708  
QY 537 YAIVLPSVQIRGSAFRQNGANNLIPWGSKVKTIGEMAFLSNRLHLDLSEQQLTRIP 596  
Db 1709 QA-----ELDEVKNLEDTVN-----QREKLVAKNS--ENDAEIDSLKEEKKALEDE 1753  
QY 597 VQAFSD--NALKE-----VLLPA-----SLKTIRE--BAFKKNHLK- 628  
Db 1754 IEKITDDNNKLSEBIDSILDRKYNALLDSKQSDVSMKEKFKQDELKVTQDALETEKKNHAET 1813  
QY 629 -----QLEVASALSHTAFNALDNDGDEQDPKNVVKVTHNSYALADGE-----HFI 675  
Db 1814 MRLKGRLEKAAEVOVRLLEALQKNLDLAQKQKAKATK-----DYRAADGKSLSNLELDDV 1869  
QY 676 VDP-DKLSSTIVDLEKILKLTGEGDYSTL-----RQTTQTPRDMTTAGKALISK 724  
Db 1870 KDQLDKAQDLDLADKXEDLTLTD-QKYKTLVKQSVFDSRIQEMQEQSL-DLEKAGRAKQK 1927

```
QY 725 SNLRQGEKQKFLQEAQFPLGRVDLDKAIKAERKALVTKKATKNGQLLERSINK----- 777
D 1928 OKQAYEKKLQELQE-----NDNDFEYKETADKRIINTLSAQKDD--LQKELEKERGLKOD 1980
QY 778 -----AVLAYNNSAIKANKVKELEKELDLTLTG-----LVGE 808
D 1981 SEKEVORLVRKQOELTKVAEYGGANVSIKVKAK-YEAEIBELATTEABDALKAKMKAEK 2039
QY 809 KGPLAQATMVQGVYLLKTLPLPEYVYIGLVNVPDKSGKLIYAL-----DM 853
D 2040 KAKTSQKKLDE---LQKT---TADYETKEASFNTGKTKQAEKLYKQYQQVRRDDETRMSSL 2093
QY 854 SDTIGGQKDAYGNPILANVDENEGYHAL 882
D 2094 EDEIKKG-TDALANKQELDKVKNQYKEL 2121

RESULT 26
AF1116
internalin proteins, probable peptidoglycan bound protein (LPXTG motif) homolog lmo0333
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AF1116
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1116
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1778 <GLA>
A:Cross-references: UNIPROT:Q8YA32; GB:NC_003210; PIDN:CAC98412.1; PID:g16409711; GSPDB:
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo0333

Query Match 3.8%; Score 195.5; DB 2; Length 1778;
Best Local Similarity 20.2%; Pred. No. 0.43;
Matches 209; Conservative 137; Mismatches 330; Indels 361; Gaps 50;

QY 44 GADYAESCKSLKINETSGPVDVTVDLFSDKRTT--PEKTKD-NLAKGPREQELKAVT 100
D 27 GADETTVSEDYAVKTAEDASATGIESETGSDDETAEPKEAKEAEASKETKEBEKAKT 86
QY 101 ENTESEKQI---TSGSLQPSQKESLSNKTVPSTSNWEICDFTKGNLTIV---GLSKSGV 154
D 87 EEPASNIKTEINTDKSQLQT---SLKAAVPAGSTY-----NSLFPDDNLAK--- 130
QY 155 EKLQSDTHVLVPSQADGTQLQIVASFATPDKTKTAIABYTSRAGENGISQLDVVGKEI 214
D 131 -KLAV---IITGNAATGNESVDSALL-----AISQLDL----- 161
QY 215 INGEVFNPSYLLKVKTIPTGYKHIGQDAFVNDKNTAEVNLPSLETISDYAFALHAKQI 274
D 162 ---SGETGND-----PTDISNIEGLQYLE--NLTSLSNEN--NISDLA-----PLK-- 201
QY 275 DLDPNLKATGELAFFDQITGKLSLPROLMRLAERAFKSNHKTIE--PRGNSLKVIGA 332
D 202 ----DLVNLVSLNLSNRNLVNLGVEDLVNLQELNVSN--KALEDISQVASFVLVKEI 255
QY 333 SPQDNDLSQMLPDGLEKIESBAFTGNPGDDHYNRVLWLTGKSNPGSLATENTYVNP 392
D 256 SAQGGCNKLTLEL-----KNPAGAVL----- 275
QY 393 KSLWQESPIDYTKWLEEDFTYQKNSVTGFSNKGLOKVKRNKQLRIP-----KQNG 444
D 276 -----PEL-----ETFYQLQENDLTNLTG--LAKLPKLKVLKYNKGNASLKSLETNG 319
```

```
QY 445 VT-----ITEIGD-----NAFRNV-DFQNKTLRKXYDL 470
D 320 ATKQLQIDASNCTDLETGLDISGLSELEMIQLSGCSKLKEITSLKNLPLNLTADSCAI 379
QY 471 BEVKLPSTIRKIGAFQSN-NLKSFEASDDLEBK-----EGAFMNN-----RIETL 517
D 380 EDLGTANLPLKQTLVLSDNENLTNITAITDLPOLKTLTLDGCGITSGTLDNLPKLEKL 439
QY 518 ELKDKLVITGDAAAPHINHIYAIVLPESVQEIERSAFRONGANNLIFMG--SKVKTLGEMA 575
D 440 DLKE-----NQITSI---SEITDLPRLSYLDVSVNLTITIGDKLPLLEWLN 484
QY 576 FLSNLEHLIDSEQKOLTEIPVQAF---SDNAKEKLVLLPASLKTIREAEKKNHLKQLEV 632
D 485 VSSNR-----LSDVSTLTNFPSLNINISNNVIRTVGKWTLPSELKEFYAQNNSISDISM 539
QY 633 ASALSHI-----AFNALDDNDGDEQDFDKVVKVTHNSYALADGEHFIIVDPDKLSSTIV-- 686
D 540 IHDMPNLRKVDASNNLTNIG--TFDN-----LPKQLSLDVHSNRTITSTSVIH 585
QY 687 DLEKILKLEGLDYSTLRQTTQTQPRDMMTAGKALLSKSNLRFQGEKQKFLQEAQFPLGRV 746
D 586 DUPSL-----ETFNAQNLNLTNIG--TMDNLPD-----LTVNLSFNKI 622
QY 747 -----DIDKAIKAERKALVTKKATKNGQLLERSINKAVLAYNNSAIKANKVYKLE 796
D 623 PSLAPIGDLPNLETLIVSDNNSYLSRLSGTMDG-----VFKLRILDIQ 664
QY 797 KELDLTLGLVEGKGPLAQATMVQGVYLLKTLPLPEYVYIGLVNVPDK-SG-----KLIYA 850
D 665 NNYLNTYCTEGNLSLSLDT-----NLTELNRNNVYDDISGLSTLSRLIY- 711
QY 851 LDMSDTIGEGQKDAVGNPILANVDEN-EGYHALA-----VATLADYEG 892
D 712 -----LNLDSNKIEDISALSNLTNLQELTLENNKIENTSALSDELN 752
QY 893 LDIKTILSKLSQTSIRQVPTAAVHRAGIFQA-----IQNAABAEQQLPKP 940
D 753 LNKLVVSKNKIIDISPVANM---VNRGAIVTASQTYTLPVLVSYQSSFTIDNPIWVD 808
QY 941 GTHSEKSSSESANSKD 957
D 809 GTLLAPSSIGNSGNYKD 825

RESULT 27
TI8353
protein p97 - Mycoplasma hyopneumoniae
C:Species: Mycoplasma hyopneumoniae
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: TI8353; T03805
R:Hsu, T.; Artushin, S.; Minion, F.C.
J. Bacteriol. 179, 1317-1323, 1997
A:Title: Cloning and functional analysis of the p97 swine cilium adhesin gene of Mycopla
A:Reference number: Z18886; MUID:97175562; PMID:9023217
A:Accession: TI8353
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1108 <HSU>
A:Cross-references: UNIPROT:Q49542; EMBL:U50901; NID:g1399524; PID:g1399525; PIDN:AAB478
R:Hsu, T.; Minion, F.C.
Gene 214, 13-23, 1998
A:Title: Molecular analysis of the p97 cilium adhesin operon of Mycoplasma hyopneumoniae
A:Reference number: Z15097; MUID:98391007; PMID:9729120
A:Accession: T03805
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1089-1108 <HS2>
A:Cross-references: EMBL:AF012905; NID:g2654768; PIDN:AAC32526.1; PID:g2654769
A:Experimental source: strain 232A
C:Genetics:
A:Genetic code: SGC3
A>Note: P97
```













Db 1503 F---GDIASVKAQYQVQVNLADSSGNOAIPTOTITNVVD---TVGPVIAKADNVSHVNT 1556  
Qy 542 PESVQIGHSAPRONGANNILPMGSKVKTLGEMAFSLNRLEHLDLSE--QKQTEIPVQA 599  
Db 1557 TKTEAEFFQADARLDVTDNN-----DDTDLIITSFAEKVNLNPKGKYEVITATDT 1608  
Qy 600 FSDNALKEVLLPASL-KTIREAFKKNHLKQLEVASA--LSHAFNALDDNDGDEQDNK 656  
Db 1609 KGNQTTKEITVQVSKDKPVITADPKISYQKIEVTEANFLSGVHAETVDELGDVKTISD 1668  
Qy 657 VVVKTHN---SYALA-----DGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLRQTTQ 709  
Db 1669 FAEKVDENKVGTVTVTLNAKDEYGNTPAEVPKVSISI--FNKIAPTNNADNKTIEAVNEL 1726  
Qy 710 QFRD--MTTAGKALLSKSNLROGEKQKPLQEAQFPLGRVOLDKAIKAERKALVTKATK- 766  
Db 1727 PSLSEIFKIEAKDYLGNKL-----KVITYPE-QTIKGNVPGYSI-----KVTTKD 1772  
Qy 767 -NQLLERGINKAVLAYNNSAIKKA-NVKRLE---KELDLL-----TGLVEG---KGP 811  
Db 1773 DSGNIAETVTLTIKDTGSPSKWTKSTKLEVSQKEPNWITFFGIKATDIDVGDVTKNI 1832  
Qy 812 LAQATMVQ---GVYLLKTPPLPEYYIIGLVYFVYDKSGKLIYALDMSDTIGEGQKDAYGN 867  
Db 1833 KVDSSSEVNLKVGTY-----PIYFTVTDALGNESKLSKSTVQIVDT-----SS 1875  
Qy 868 PILVUDEDNEGHALAVATLADYGL-DIKT-ILNS--KLSQTSIRQVPTAAYHRAGIF 923  
Db 1876 PELTIDKLRISYP--TGKTSIDKQFLQDGTGKVTNSYGTVKVTTNLSKI--VDWNAKAGY 1931  
Qy 924 OAIONA-----AAAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKNRRGHSAILPR 977  
Db 1932 KVTVTATNSGGVAEKTIILL-----TVNSDSSFIAPVPSKD-----DNKNPKAKNIPK 1979  
Qy 978 TG 979  
Db 1980 TG 1981

RESULT 35  
T18354  
adhesin - Mycoplasma hyopneumoniae  
C:Species: Mycoplasma hyopneumoniae  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18354  
R:Wilton, J.B.; Scarman, A.L.; Walker, M.J.; Djordjevic, S.P.  
Microbiology 144, 1931-1943, 1998  
A:Title: Reiterated repeat region variability in the ciliary adhesin gene of Mycoplasma  
A:Reference number: Z18887, MUID:98361039, PMID:9695926  
A:Accession: T18354  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1092 <WIL>  
A:Cross-references: UNIPROT:O07132; EMBL:AF001398; NID:g2150123; PID:g2150124; PIDN:AAC3  
C:Genetics:  
A:Genetic code: SGC3

Query Match 3.7%; Score 186.5; DB 2; Length 1092;  
Best Local Similarity 18.7%; Pred. No. 0.54;  
Matches 203; Conservative 159; Mismatches 368; Indels 353; Gaps 48;

Qy 20 HNGEVSFLKEPILKQTA---SSSISGADYAESSGSKLKNETSGPVDVDTDLFSDK 76  
Db 142 HNGDI---AKSDIYEQTAVAFKQSNLLVAEFNPSLKKITEKLNQOINLSKTN-FADE 197  
Qy 77 RTTPEKIKONLAGPREQELKAVTENTESEKQITSGSQLEQSKESISLKNTPVSTSNWEI 136  
Db 198 KTSQKDPSTLRAIDFYDLN-TARNAE----- 224  
Qy 137 CDPITKNTLVGLSKGVKLSQTDHLVLPQQAQDGTQIQAQVAFPTPKKTAIAYETS 196  
Db 225 -DLIDIKLANYFPVLKNIINLNNAENKLPN-----NLGNIFEFSAKDSST--NQYVS 275

Qy 197 RAGENGBI---SOLDVDGKEIINEGE---VFNSYLLKKVVTIPTGYKHIGODAFYDN--- 246  
Db 276 IQNQIPSLFLKADLISQSAREILASDEVPQVIN--ILRLMKDNSSYFLNPFEDFVNNLT 333  
Qy 247 KNIAEVNIPESLETISDYAF-----AHLALKQIDLPDNLKAIGELAFPDN 291  
Db 334 KMQKEDLNAGQNLQSAVEFLADIKSGFFPGDKRSSHTKABEISNLLNKKENIYDFGKYNG 393  
Qy 292 QITGKLSLPLQMLRLAERAPKSNHKTIEFRGNSLKVIGEASFQDNDSLQMLPQGLEKI 351  
Db 394 KFNDRLNSP-----NLEY---SLDA-ASASLDDKDKXSILIPYRLE-I 431  
Qy 352 ESEAFPTGNPGDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSL-----WQESPEID--- 403  
Db 432 KDKFF---ADLLYPD-----TKDNILVKEGILKLTGFKGPKLDLNI 471  
Qy 404 -----YTKWLEEDFTYQKNQSVTSPNSKGLQV-----KRNKNLUIPIKOHNGV 445  
Db 472 NQOIPKTEVLPFFKEGKEEQAKLDYGNILNPYNTQAKVEVEALFKGNKQBIYQALDGN 531  
Qy 446 TITELGDNVAFNV-----DPQNTLAKYDLEEVKL-----PSTIRKIG 483  
Db 532 YAYEFG--APKSVLNSWTGKIQHPEKADIORFTRH---LEQVYKLGNSVNLQPOPTKEQV 586  
Qy 484 AFAFOSNNLKSFASDDLEIEKEGAFMNRITLLELKDVLVTIGDAAFHINHIIYAI--- 539  
Db 587 ISSLSKNFF-----FANGHOVASYQDILLTKDKL-TVLSTLDLAKKWLGNR 634  
Qy 540 -VLPESVOEIGRSAPFRONGANNILPMGSK-----VKTGEMAFSLNRLEHLDLSEQKQ 591  
Db 635 AQFPKEVQYTKDIFAE--ADKLFEGCKKDPYVQIKIEIHQLSP--NILARNDVIK--- 687  
Qy 592 LTPVQAFPSNALKEVLLPASLKTIRREAFKKNHLKQLEVASASLASHAFNALDNDGDE 651  
Db 688 -----SDGFYGVLLLPQSVKT---ELEGKNEAQ-----IFEAL----- 717  
Qy 652 QPDNKKVVVTHHNSVALADGEHFIVDPKLSTIVDLEKILKLEGLDYSTLRQTTQTOF 711  
Db 718 -----KYSLEIENSAF-----KTIILD-----KNLEGTDFKT-----F 746  
Qy 712 RDMTTAGKALLSKSNLROGEKQKPLQEAQF--FLGRVLDLKAIAKAEKALVTKATKNGQ 769  
Db 747 GDFL-----KAFLLKAAQFNFPAPWAKLDDNQLQYSFEAIKGETTKEG- 789  
Qy 770 LLIERSINKAVLAYNNSAIKKNVKKLEKELD-LLTGLV-----EKGGLAQATWQOG 820  
Db 790 -----KREBVDKVKELDNKIKGILPQPPPAKPEAAKPVAAKPEAAK 831  
Qy 821 VYLLKTPPLPEYYIIGLVYFVYDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDENEG-- 878  
Db 832 PETTKPVAAPPE-----AAKPVAA-----KPVAAKPVATNTNTNTGFS 869  
Qy 879 -----YHALAVATLADY---EGLDIKT-----ILNSKLSQTSIRQVPTAAYH 918  
Db 870 LTNKPEKEDYFPWAFSYKLEYTDENKLSKTPEINVFLVHQSEYEEQKIKELDKTVLN 929  
Qy 919 RAGIQAIONAAAEAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKNRRGHSAILPRT 978  
Db 930 LQYQOEKVKVTSEYQKLSHPMTTEGSPNOGKKAEGAPNOGKKAEGAPSGQKKAEGAPNQ 989  
Qy 979 GSK 981  
Db 990 GKK 992

RESULT 36  
BHEYDL  
RAD50 protein - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein N0872; protein YNL250W  
C:Species: Saccharomyces cerevisiae  
C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 09-Jul-2004  
C:Accession: S05808; 563223

Qy	546	QEIGRSAPFRONGANNLIFPMGSKVKTIGEMAFPLNRLHLDLSEQKQLTEIPVQAQPSNAL	605
Db	971	D-----FEARGFDEL-----QTTIKELEL--NKAQMLEKEQLDKSNEV-----NEE	1011
Qy	606	KEVLPLASLTIRREAFKKHNLKOLEVASALSASHIAFNALDNDGDGFQDNKVVVVKTHNS	665
Db	1012	KRKLADSN-----NEEKVNKNLEIEUKSQLOHI-----EESURL-----RTRPEKLSSENAGKL--	1082
Qy	666	YALADGEGHFIVDPDKLSTTVDLLEKILKIEGLDYSLRTQTOTQFRDMTT--AGKALLS	723
Db	1045	ISRLDVQNAERDKYO-----EESURL-----RTRPEKLSSENAGKL--	1082
Qy	724	KSNLRQEBKQKFQEAQFFGRVDLDKAIKAELAVTKK-ATYNGQLLERSINKAVLAY	782
Db	1083	-GENMKOLQONQIDSLTHQLRTDYKDIEKNYHEWVELQTRSFTVDDIDIVYSKALDLSAIMKY	1141
Qy	783	NNSAIKXANKVRLEKEL-----DILTLGVBEKGKPLAQATMVGQVLLK	825
Db	1142	H--GLKMQDIRRIIDELWKTYSGTDITIKIRSDEVSSTVKGSYNRYVMYK-----	1193
Qy	826	TPLPPEYYIGLVYFDKSGK-----LIYALDSMTTIBSG-QKDAYGNPILNVDE	874
Db	1194	-----QDVELDMRGCSAGQKVLAIIIRLASLSETFGANGCVIALDEPTTNLDE	1242
Qy	875	DNEGYPHALAVATLADYEGLDIKTIILNSKLSQ	905
Db	1243	EN-----IESLAK-----SLHNIIINRRHQ	1262
 RESULT 37 S22452			
surface exclusion protein seal precursor - Enterococcus faecalis plasmid pAdI			
C:Species: Enterococcus faecalis			
C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Jul-2004			
C:Accession: S22452; S22069			
R:Weidlich, G.; Wirth, R.; Galli, D.			
Mol. Gen. Genet. 233, 161-168, 1992			
A>Title: Sex pheromone plasmid pAdI-encoded surface exclusion protein of Enterococcus faecalis			
A:Reference number: S22451; MUID:S2293110; PMID:1603060			
A:Accession: S22452			
A:Molecule type: DNA			
A:Residues: 1-890 <WEI>			
A:Cross-references: UNIPROT:Q01891; EMBL:X62658; NID:g49019; PIDN:CAA44526.1; P			
C:Genetics:			
A:Gene: seal			
A:Genome: plasmid			
 Query Match 3.7%; Score 185.5; DB 2; Length 890; Best Local Similarity 19.9%; Pred.No. 0.45; Matches 209; Conservative 162; Mismatches 410; Indels 267; Gaps 49;			
Qy	7	TVALTLTVSVVTHNQVFSLVKEPIKQTPASSISSGADYABSSGSKLKINETSGPVD	66
Db	45	TVKATQTTEAITEKQQQVT-EKQAVDQKQ-----QVADTAKEK-----	84
Qy	67	DTVTDLPSDKRTPPEKIKNLAGPREQELKAVTEN-----TESEKQI--TSGSOLEQSKE	120
Db	85	DAIDSQSVKDAQVVDQNKDAL-----DQSOQAVTDQAAVVDQAAKVVDEATPSAIEAKE	139
Qy	121	SLSLNKTVPTSPSNWEICD---FITKGNLTGLSKSGV-EKLSOTDHLVLPQADGGQLI	176
Db	140	QV-----ATDQVDDQKQVDDQAGTDVNOQAAVVEEKAKETN-----	177
Qy	177	QVASFAPTPDKTAIAEYTSRA-----GENGEISQLDVDGCKEIINEGEVFNLYLLKKVTI	231
Db	178	-AAKVQNDKQQAATAKQEQAKLEELAKNAEAKAEKAEQAELANKQ--KEEAK	234
Qy	232	PTGYKHGTQDAFVDNKNIAEVNIPESLFTISDYFAFHALLKQIDLPNLKAIGELAFPDN	291
Db	235	AKDQKTDDQAVADQQSV----VTTSOBKVAD-AKADTAARQADLTISK-----EN	279
Qy	292	OITGKLSLPROMELASRAFKSNHHKTIETPRGNSLKVIGEASFODNLSOIMLPDGLEKI	351

Db 280 ALKDKQAATKQ-----AQNLTDSSEELKKGKGLNL-----PANFTF-DVYKKLITEQKQAM 330  
Qy 352 ESEA-----FTGNPGDDHNNRVVLTWTKSGNPNPSGLATENTVYVNPDKSLQWSEFIDYT 405  
Db 331 EKXALALANKVPENQADAQKVTMI-----NVKNP-----TEK-----OKQMSDYV 372  
Qy 406 KWLIEDFTYQNSVTFGNSKGLQKVK-RNKNLEIPKQHNGVTTITBIGDNAPFNVDQNK 464  
Db 373 VGLINDVR-----EKLGLQKLSIQAOMKF-----NGANNLFMSGKVKTKLCEMAFL 577  
Qy 465 LRKYDLEEVKLPSTIRKIGAFQSNULKSFPASDDLEBEIKGAFMNNRIETLEKOKLV 524  
Db 409 EFDHVNAINR-----AAKENGKFKFPQNFYENLSMG-----RFTTQEGKVMY 453  
Qy 525 TIGDAAFHNIYAIVLPESVQIEGKSAFRQ-----NGANNLFMSGKVKTKLCEMAFL 577  
Db 454 DPKKARN-----ALVSMLANDGHSYSHLDSLLDANETV-----MAVISGDLNDI 500  
Qy 578 SNRLEHDLSEKQKLTETIPVQAFSDNALKEVLLPASLKTIREEAFKGNHLKQLEVASALS 637  
Db 501 SAKI-HIISYNTQKLVFA-----NTVEEGTAP-----VFKSKETLQKEVATNQE 543  
Qy 638 HIAFVALDDNDGDE-OFNKVUVKTHNSYALADGEHFIVDDPKLSSTIVDLKILKLE 696  
Db 544 KLATAQAESDAQARSASQAALNTAKTQATAEKELSV-----HKATLANLQEV----- 593  
Qy 697 GLDYSTLQTTQTFQDRMTTAGKALLSKNSLRQGEKQKFLQEAQPFPLGRVLDLKAIAKAE 756  
Db 594 -----ATKSTTNYEKKVQTAA-----AETNLQQTQKQ-----LATINELI--QN 631  
Qy 757 KALVTKKA-TK--NQOLLERSINKAVLAYNNSAIK--KANVKRLSEKDLTLTGLVEGKGP 811  
Db 632 RAVVLEKATKVADAQAIEQTSK-VLKEQEAQAEENTLNSLKEVLDLAKENLNQK-Q 689  
Qy 812 LAQATWOGVYLLKTPLPPEYIGLVYFDKSGKLIIVALDMSFTDGSKDQDAYGNPILN 871  
Db 690 VALKSTSTRSLARLENAQPTYE-----KAVNELKAAEAAYVQAEAVENSLKS 736  
Qy 872 VDQDNQGYHALAVATLADYEGLDIKTLNLSKLSQTSIRQVPTAYHRAQIFQATONAAA 931  
Db 737 LEHLKE-----QQAVALATLQQAEDLSNAKLEQQYQGLRDLQEAQAEQRQEAQV 793  
Qy 932 EAKQLLPKPGTHSEKSSSESANSKDRGLQ-----SNPKTNRGHS----- 972  
Db 794 KEQQLREAKRQQLMVASATSADTKPGIQLQSLFSKQEQKPAQTLTHSEPRKTKQVAKA 853  
Qy 973 -AIIPTGSGSGFVYGIIG-YTSVALLS 998  
Db 854 QESLPHTEGQKSIWLTIFGLFMAVGAIS 881

## RESULT 38

G69708

Chromosome segregation SMC protein - Bacillus subtilis

N;Alternate names: minichromosome stabilizing protein SMC

C;Species: Bacillus subtilis

C;Date: 05-Dec-1997 #sequence revision 05-Dec-1997 #text\_change 09-Jul-2004

C;Accession: G69708; JC4819; FC4029

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gall  
isch, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidin, A.; Lardinois  
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: G69708  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1186 <KUN>  
A;Cross-references: UNIPROT:P51834; GB:Z99112; GB:AL009126; NID:G2633902; PIDN:CABI3467.  
A;Experimental source: strain 168  
R;Oguro, A.; Kakeshita, H.; Takamatsu, H.; Nakamura, K.; Yamane, K.  
Gene 172, 17-24, 1996  
A;Title: The effect of Srb, a homologue of the mammalian SRP receptor alpha-subunit, on  
A;Reference number: JC4819; MUID:96257247; PMID:8654983  
A;Accession: JC4819  
A;Molecule type: DNA  
A;Residues: 1-49, 'G', 51-161, 'G', 163-174, 'E', 176-177, 'G', 179-191, 'G', 193-227, 'P', 229-235,  
74, 'P', 476-493, 'D', 495-514, 'D', 516-541, 'V', 543-545, 'P', 547-585, 'SKPLRNSGPAFIISF', 601-622,  
'F', 741-1186 <OGU>  
A;Cross-references: DDBJ:D64116; NID:G1389548; PIDN:BAAL0977.1; PID:G1237015  
R;Oguro, A.; Kakeshita, H.; Honda, K.; Takamatsu, H.; Nakamura, K.; Yamane, K.  
DNA Res. 2, 95-100, 1995  
A;Title: srb: a Bacillus subtilis gene encoding a homologue of the alpha-subunit of the  
A;Reference number: JC4093; MUID:96093930; PMID:7584053  
A;Accession: PC4029  
A;Molecule type: DNA  
A;Residues: 1171-1186 <OG2>  
A;Cross-references: DDBJ:D49781; NID:G1237017; PIDN:BAAO8615.1; PID:G1237018  
C;Genetics:  
A;Gene: smc  
C;Superfamily: chromosome segregation protein SMC1

Query Match 3.7%; Score 185.5; DB 2; Length 1186;  
Best Local Similarity 20.1%; Pred. No. 0.68;  
Matches 211; Conservative 177; Mismatches 403; Indels 259; Gaps 53;  
Qy 2 KKHUKTVALTITTVSV-VTHNQEVSLVKEPILKQTAQSSISGADYAESGKS--KUKI 58  
Db 219 KKELEHVEIALTAYDIEKLHGK--WSTLKEVKVMAKE-----BELAESAISAKAKI 269  
Qy 59 NETSGPV---DDTVTDLPSDKRTTPEKIK--DNLAKGPREOELKAVTENTESEKOITSGS 113  
Db 270 EDTDKKQALDESQVNELOQVLLVTSELEKLEKGRKEVLEKKNKAVNQOEQLSEAIQVQ 329  
Qy 114 QLEQ-SKESLSNKTVPSTSNWEICDF---ITKGNLTVLGSKSGV-EKLSQ--TDHL-VL 165  
Db 330 QKETVLKEELSKQEAQVETLQAEVKQLRAQVKQKQALSLHNEVVEKIEQLKSDYFELL 389  
Qy 166 PSQAADGTQLIQVASFAPTPDKTAIABYTSRASENGE---ISQLDVQKKEIINEGEVFN 222  
Db 390 NSQASIRNEL-----QLLDDQMSQSAVTQLRLADNNNEKHLQERHDI SARKAACETEF-- 441  
Qy 223 SYLLKKTPTPTGYKHIGQDAFVDNKNIAEVN-----LPSLETI-SDYA 265  
Db 442 ARIEQEIHSGVGAQVDRDMQTKYEQKQKRNESALYQAYQVQQAARSKKOMLETMQGDFS 501  
Qy 266 FAHLALQKIDLPDLNKAIGELAFFDNQITGKLSLPRQLMRLAERAF--KSNHI----- 316  
Db 502 GFYQGVKEV-----LKAKERLGGIRGAVLELITSEKQYETAIEIAGLSAQAHVVTDDQS 556  
Qy 317 --KTIETP-RGNS-----LKVTEASFPONDLSQLMLPDGLEKIESEAFTEGPDH 365  
Db 557 ARKAIQVLKQNSFGFRATPLPLSRDQLQSRDAETAAHSSSFLGVASELVTFDDA---Y 613  
Qy 366 NNRV-----VLWTKSGNPSGLATENTVYVNPDKSLQWSEFIDYTKWLEEDFTYKXSV 419  
Db 614 RSVIQLNLGTVLLITFEDLKGANELA-----KLLGHRYRIVT-----LEGDVNPGSGM 660  
Qy 420 TGFSNKGLQKVKRNKNLEIPKQHNGVTTITIGDNAPFNVDQNK-TLKYDLEEVKLPST 478  
Db 661 TG-----GAVKKNNSLLGRS-----ELEDVTKLAEEMEETALLEQEVKTLK---- 704  
Qy 479 IRKIGAFQAFQSNULKSFPAS--DDLEETKEGAFMNNRIETLEKOKLVITIGDAAPHIN-HI 536  
Db 705 -----HSIQDMEKKLADLRETGEGL-----RLKQQDVKGQLVQELQVAKENINHL 749







```
Db 610 -----ELKAIEGSAVAVLESGVRGK-----AVATVIVSGTSLVSGDFFAETAP 657
Qy 678 -----PKLSSTIVDLE-KILKLEGLDYSTLRTOTQFDRMTTAGKALLSKSN----- 726
Db 658 GKVRTWTDQGSIGNLKPMSVALITGLS-----EVPPAGSVLIGVENDSIA 704
Qy 727 -LRQGEKQKFLQBAQFFLGRVDLKAIAKAEXA-----LVTKKATON-----G 768
Db 705 RLQAKGRATYLRQ-----KALSKSTKVSFDELSEMVANKELKNI PVVIKADTQG 753
Qy 769 QLLERSINKAVLAYNNSAIKKANVR-----LEKELDLL-----TGLVEG 808
Db 754 SL-PAIKNSLLELNEEVAIOVHSVGGITENDLSVSSSEHAVILGFNPRTGNVN 811
Qy 809 KGPLAQATWQGVYLLKTPLPPEYVIGLVYFDKSGKLIIVALDMSDTIGBGQKDAYG-- 866
Db 812 KA-----KEYNVSIKTY-----TVIVAL-----IEEMRSELLGLM 841
Qy 867 NPILNVD-----EDNEGYHALAVATLA 888
Db 842 SPIIEEHTGQAEVRETFNPVKGTIA 868

RESULT 42
P96673
Hypothetical protein F13011.30 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: F96673
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizlar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141, MUID:21016719, PMID:11130712
A:Accession: F96673
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1313 <STO>
A:Cross-references: UNIPROT:Q9XIP6; GB:AB005173; NID:G5042434; PIDN:AAD38273.1; GSPDB:GN
C:Genetics:
A:Gene: F13011.30
A:Map position: 1

Query Match 3.6%; Score 182.5; DB 2; Length 1313;
Best Local Similarity 20.2%; Pred. No. 1.1;
Matches 211; Conservative 171; Mismatches 380; Indels 281; Gaps 51;

Qy 31 PILKQTOASSISGADYAESSCKSLKINETSGPVDVTDLPFSKRTTPEKIKDNLAKG 90
Db 22 PRLSKLSASKDS-----NSAPVHSLRVKG-----TELQTLNQIQEDLKA 65
Qy 91 PREQEL-----KAVTENTESEKQITSGSQLESLSLNKTPSTGNWEICDPIITKGN 144
Db 66 DEQIELLKDKAKAIDDLKESEKLVBEAN-EKLEALAAQR--AESPEVEKF----- 116
Qy 145 TLVLGSLSGVEKLSTQD-----HLVLPQADGTQLIQVA-----SPAFTPDKK- 188
Db 117 RAVELEQAGLEAVQKDVTSKNELESIRSQHALDISALLSTTEELQVRKHLSMTADAKN 176
Qy 169 -----TATAEYTSR-----ACENGEISOL--DVDGKEIINEGEVFNLSLLKVTIP 232
Db 177 KALSHAEBATKIAEIAHAKAEIISLGLKALLGSKBEKEAI-EGNEIVSLKSKSEIELL 235
Qy 233 TGYKHIGQDAFVNDKNIAEVLNIPESLETISDYAFALHAKQIDLPDNLKAIIGELAPFQNG 292
Db 236 RGE-----LEKVSILESSLKE-----QEGLEQLKVDLEAA----- 266
```

```
Qy 293 ITGKLSLPRQLMLAEAFKSNHIKTIEFRGNSLKVIGEASFPQDNDLSQMLPDGLEKIE 352
Db 267 -----KMAEECTNS-----SVEEMKNKVHEL-EKEVEESNRSKSSASESMESVM 309
Qy 353 SEAFGNPGDDHYNNRVVLTGSKGNPSGLATENTYVNPDKSIWQESPEIDYTKWLBEDF 412
Db 310 KQLAE-----LNHVLUHETKS-----DNAAQKEKIEL-----LEXTI 340
Qy 413 TYQNSVTGFSNK-GLQKVRENK-----NLEIPKQHGVTIITEIGN-----AFRNV 458
Db 341 EAQRTDLEEGROVCIKAEKLENLVSLEISQEEK-----TRALDNEKAATSN 396
Qy 459 D--FQNTLRKYDLBEVKLPSTIRKIG-----APAFQSNLKSFPASDDL-----FEIKEGA 508
Db 397 QMLDQRTLELSIELERCKVEEKSCKDMESITLALQEAESTESSEAKATLLVCBELKN-- 454
Qy 509 FMNRIETLELKDCLVT-----IGDAFHNHIVAVILPESVO-----EIGRSAPRQNGAN 559
Db 455 -CESQVDSLKLASKETHEKYMLEDARNEIDSLKSTV--DSIQNEFENSKAGHEQ---K 508
Qy 560 NLIFPMGSKVKTGEMAFLSNRLEHLDSQKQLTEIPVQAFSDNALKEVLLPASLKTIRE 619
Db 509 ELHLMGCVKKS-----EEENSQSEVSRV-----NLLKESEBDACARK-EE 550
Qy 620 EAPKKNHLKQLEVASALSHTAFNALDDNDGDEQFNDKVVVKTHTNSYALADGEHFIVDPD 679
Db 551 EASLKNLKVAB-----GE-----VKYLOETLGEAKESM----- 580
Qy 680 KLSFTIVDEKILKLEGLDYSTLRQ---TTOTQFDRMTTAGKALLSKSNLRQGEKQKFL 736
Db 581 KLESLLDKEDLKNVTA-EISSLRWEGSVLEKIEBSKVKESLVK---ETKLSQIT 635
Qy 737 QEAQFFILGRVDLDKAIKAERKALVTKKATKNGQLLERSINKAVLAYNNSAIKKANVRKLE 796
Db 636 QEAEELKGR-----EAAHMKQIEELSTANASLVDEATKLQSIQVQESDELKEKEAGYLK 688
Qy 797 K--EIDL-----LTGLVSGKPLAQATWQGVYLLKTPLPPEYVIGLVNVPDK 843
Db 689 KIEBSVANESLADNVTDLQSIQVQESKDLKEREV---AYLKK---IEELSVANESLVDK 741
Qy 844 SGKLIYALDMSDTIGEGQKDAYGNPILNVDSDNGYHALAVATLADYEGL--DIKTILNS 901
Db 742 ETKLQH-IDQAEHLRGREASHLKIEELSKENEN-----LVDNVANQNIABESKDLRER 796
Qy 902 KLSQTSIRQVPTAAYHRAGIFQAIQNAABAEQLPKPGT-----HSEKSSSSSESANSKD 957
Db 797 EVAYLKKIDELSTANGTLADNVNLTQNISEENKELRERETTLKKABELSELNESLVDKA 856
Qy 958 RGLQSNPKTN---RGRHSAILPR 977
Db 857 SKLQTVVQENBELRERETAYLKK 879
```

## RESULT 43

T43527

sp8 protein - fission yeast (Schizosaccharomyces pombe) (fragment)

C:Species: Schizosaccharomyces pombe

C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 04-Feb-2000

C:Accession: T43527

R:Jimenez, M.; Goday, C.

submitted to the EMBL Data Library, June 1997

A:Description: Schizosaccharomyces pombe myosin-like.

A:Reference number: 222545

A:Accession: T43527

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1173 &lt;JIM&gt;

A:Cross-references: EMBL:AF010473; PIDN:AAB65416.1

A:Experimental source: strain 972h(-)

A:Gene: sp8

```
Query Match      3.6%; Score 182; DB 2; Length 1173;
Best Local Similarity 20.2%; Pred. No. 0.98;
Matches 199; Conservative 163; Mismatches 326; Indels 296; Gaps 48;

QY 33 LKQTOASSISGADYA-----ESGSKKLKINETSGPVD-----66
Db 281 LHSLOKSEKSGEYSRIKTLMLKQSI SEDNRKLLDNQOMWEIKQLBNGVIELEKOR 340
QY 67 -DVTDLDFSDKRTTPBKIDNIAKGPREQELKAVTENTESE-----KQITSGSQLE- 116
Db 341 FSTLEAKFTQOKNTYSEREALESLSL-SDLOKSHTSLESQYNSLRNIEQLQAASKLAE 399
QY 117 -----QSKSLSLKNKTVTPSTSNWEI-----CDFITK-QNTLV 147
Db 400 EMVERVKTEBYDEVLQTSSELSKRNHLKITSLEQRIVILQDEIATSSSLRCENITKQSETRV 459
QY 148 GL-----SKSGVEKLSQTH---LVLPQNAQDT-----QLIQVAS 180
Db 460 ALLEENKHNLSHRNAEKHLEK--ENDYKQQLLVTEDLRKTRDEYKELLRHAD 517
QY 181 FAPTDPKTAIAEYTSRAGENGESIQDVGKEINEGEVVF-----NSYLLKVV 229
Db 518 ARSTLQK--LRDYTKALEQVEDLNEKIALKAGINESQFPPISEKEDPLRQEVVYVKK- 573
QY 230 TIPTGYKHIGQDAFV-----DNKNIAEYNLPE-SLETISDYAPAHALAKQIDLPNLKA 282
Db 574 -----QNAMLLTQLQSSNLNFAEITSPSPDLDSV-----MKLGLSLQNHVXR 616
QY 283 IGLAFDPNQITGKLSPLQMLRAERAPKSNHIKTIIEFRGNSLKVIGASFDNDLSOL 342
Db 617 ISK-----EMEI--ISCQQLFLENKCLK--RTVE--SSNRVIAD-----651
QY 343 MLPDGLKEIESEAFTCNPGD-DHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLMQESPE 401
Db 652 -LQRGTEKDVSTSESGERSNYLNVALLESNKs-----LRENLEARN-EEVITELREK 705
QY 402 IDYTKWLEEDFTYQKNSVTG--FSNKGLOKVRKNKLEIPKQHNGVTITEIGDNAPRNV 459
Db 706 IETLKTDLANFLRNKLEQLSOLQTEKAAVKCLSENSNEY-KRHNQBIILLSNSTSTSD 764
QY 460 ----FQNKTRKYDL-BEVKLPSTIRKIGAPAFQSNLKSFEASDDLEBEIKGAFMNRNIE 515
Db 765 ASRLKNELVSKENLIBELN-----QBIGHLKSELETVKS--KSEDLLE--NERAQNOSKIE 815
QY 516 TLRLKDKLVITGDAAHINHIAVLPEVSQVEIGRSAPFQNGANNLI FNGSKVKTILGEWA 575
Db 816 QLELKNKLA--AAWRTK--YEQVVKNSLEK--HNQIRQQLSQKTSSELEAKVAECHOJN 868
QY 576 FLSNR-----LEHLD-----LSEKQKLTET-----PVQAFSDN 603
Db 869 EQLNKPSATPTATQSEPTSVLSLEEFNSTKEELSSSTQKLSLEIMDILNTTKEELEKVRQN 928
QY 604 ALKEVLLPASLTIRREAFKKNHLKQLEVASALSHIAFNALDDNDGDEGFNKNVVKVTHH 663
Db 929 SNKSEKTSKDTPIPEEERERKKVMQOEVLRLRSRIA-----KELQKNELLARK---976
QY 664 NSVALADG---EHFIVDPDKLSSTIV-----DLEKILKLIBEL-----DYS 701
Db 977 QNOVLQDQVKALQETVVSSEEAESASVHADTKOLENLKKEETEMLSVTFQVIFNESISDFS 1036
QY 702 TL-----RQTQTQFRDMTMTAGKALLSKSLRQCKQKFLQ 737
Db 1037 TSTADPTFTTVQKEWEKREILQKQVEEQASHQKQLDNIRKELENRNKLKLSMLEKNLA 1096
QY 738 EQAFFLGRVDLKAIAKASKALVTKATKNGOLLERSINKAVLYNNSAIKKANVVRLEK 797
Db 1097 RV-----RALEQSKKKDSPAILLSLEASKN-----TDSNK-----SNSEVPAAQVK--EK 1139
QY 798 ELDDLTLGLVEGKPLAQATWQGV 821
Db 1140 KLIATKTHSVDTNTPPKRSSSDAGM 1163
```

```
RESULT 44
AE1717
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AE1717
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Bloeker,
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitouran, A.; M
ok, C.; Schlueter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669.
A:Accession: AE1717
A:Status: preliminary
A:Molecule type: DNA
A:Accession: 1-1622 <GLA>
A:Cross-references: UNIPROT:Q929J4; GB:AL592022; PIDN:CAC97509.1; PID:gl6141793; GSPDB:
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin2281

Query Match      3.6%; Score 182; DB 2; Length 1622;
Best Local Similarity 18.9%; Pred. No. 1.6;
Matches 240; Conservative 165; Mismatches 436; Indels 430; Gaps 58;

QY 2 KKHKTVALTLTVSVVTHNQEVFSLVK-EPILKQTOASSISGADYAESGSKSKLINE 60
Db 64 KKTGETKSVISDGTIVEN-----SLSEGEYIVKKAAPGT-----LDEQTYNV 110
QY 61 TSGPVDVDTDLFSDKRTTPEKIDNIAKGP-----REQELKAV-TENTSEKQITSG 112
Db 111 TLADKDEVITSI-----STKDEAKTPIVTEQPSKKGNLKAVIDIADNIFTAVKENG 161
QY 113 SOLEQSKES-----LSLN-----KTV-----PST- 131
Db 162 TGNELSTTNRINKGAVVLKNNFTFSGKNYKAGDTFKVLPTFNFGTTLNLTGDFLPSFE 221
QY 132 SNWE-----ICDFTKGNLVLGSKSGVEKLSQTD-----H 162
Db 222 AKWDLNASTRELITITFLKDGVOEGNYDIELS-TALKSFSETDKTRQEAVENTAGNTIYQ 280
QY 163 LVLPQQAADGTQ-LIQVASFAFTPKDKTAIAEY--TSRAGENGESIQDLDV--DGKEINE 217
Db 281 LEIIEVDKATQVMLEATPGKLPKATVDARFNLTKESNELGELRLSDTAYGSTIINR 340
QY 218 G--EYFNSYLLKCKVTIPTGYKHIGQDAFVDNKNIAEVLNPESETISDYAFALHALKQID 275
Db 341 NSIKVYSTDISAKGT-----FIGSKQL--LNEGTDYELIYAPSGLTIK---381
QY 276 LPDNLKATGELAFPDNQI-----TGKLSLPRQLMRLAEAFKSNHIKTIE-----FRGNSL 326
Db 382 LKEGLKAKGYQVTYERSIDKNTSLTTLGTSATTVSSGSMLSNGSMITISVTIKAYDHLTK 441
QY 327 KVIGEASFQ-----DNDLSQLMLPDGLE-----KIESEAFNGP---360
Db 442 KAVNPVTQCIDWTNINVDLANLTPDTVLKDVLTDDNVSIVYADSLKIKQVTFNBSGEV 501
QY 361 --GDHYNNRVVLWTKSGKNPSG-----381
Db 502 VIGEDASND---WTVSTISDNGSFNNYKTKDKKAYQVYTYSTKLTDFSPRIKNEVTDE 557
QY 382 ---LATENTYVNPDKSLMQESPEIDY-----TKW-----LEEDFTYQKNSV 419
Db 558 KGVKATENPDFKPD-LLKKEAGEIDYNNYNTMDTWTITVNSEGINMQINIVDEFSTGVKSL 616
QY 420 TGF-----SNKGLQKVRKNKLEIPKQ--HNGVTITEIGDNAPRNVFQNKTLKPYDL 471
Db 617 ISYSVYVPSDSGYNILTEGRDFTIQKDVTPDGFKIKLIGNATTATDNKIVVAKTKIDLT 676
QY 472 EVKLPSTIRKIGAPAFQSNLKSFEASDDLEBEIKGAFMNRNRIETLELKDCLVLTIGDAAF 531
```

Db	677	DC--ANTLNKASFYFDGSLTQYSEIKAEATPETSILAN-----GGKVGKWNPATG	727
Qy	532	HINHIYAIVLPESVQIEGRSAPRONGANNLIPMGSKVKTLCGEMAFSLNRLEHLDLSEQKQ	591
Db	728	EINWIV-----SVNMGKK-----YKLVLI---DDEFLDGTTFVEGSLQTRNVNSSE	772
Qy	592	LTEIPVQAFSONALKEVLLPASLTIRBEAFKKNHLKOLEVASALSHIAFNALD-----	645
Db	773	LTDLSI-----PLEIKGTLAQVGDNYPYTKIDTSANKIHLFEGNLDINRVFV	819
Qy	646	-----DNDGDEQ-----PDNKVVVKTTHNSVALADG---	671
Db	820	KYKTKPKONWFPQWNNKAIUSDNGADEQIYETKEFAFLQNEVIKVAIGNIDSTFGNKVN	879
Qy	672	---EHPYVDPDK-LSSTIVD-----LEKILKLI-----EGLDYSLRQ	705
Db	880	MNMLLNITPERTLSNPVITNRLEQNTGAQFIKNSFQVINTKTNEPINEENYDITFEGN	939
Qy	706	TTQTOFRDMTTAGK-----ALLSK--SNLRQGEKQKFLQEAQFF-----	742
Db	940	FTTIQKNTAMAPIKVNYSITISLLSGPISNETTVEADFSNVPMPFFKKRNVAVSPVTV	999
Qy	743	-----LGRVDLDKAIAKAKA--LVTTKATKQGQLLERSINKAVLAINNSA	786
Db	1000	GTGSGIATIGTIKITKVDDEDTTKLEGAKFQLYTLDGESKQBEIT-----TNS-	1048
Qy	787	IKKANVKLEKELDLTLGLVEGKGLAQATWQGVLYLKTPLPBYVYIGLVNVPDKSGK	846
Db	1049	-----EGBI-LLDGTSQSKYKLVETKAPEGNYI-----SDEYKEGKEITVSSGE	1092
Qy	847	LIYALDMSDTIGEGOKDAVGNPILNVDENEGHALAVA--TLADYEGLDIKTILNSKLS	904
Db	1093	-----ELLTINKVANK--GNVIL--TKDASHEVLADAEFLQNAATGSKLEKLTDDAS	1144
Qy	905	QUTSIROVPTAAHYHRAGIFQAIONAAAEQLPKPGTHSEKSSSSSESANSKORGLQSNP	964
Db	1145	GNIETDL-----APGDYKLIETKAPTGYQLDATPVNFTIDFNQSEAAKVS-----	1190
Qy	965	KINRGHSAIL	975
Db	1191	KNTAKTGTVM	1201
RESULT 45			
AE1449			
Hypothetical protein lin0132 [imported] - Listeria innocua (strain Clip11262)			
C:Species: Listeria innocua			
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004			
C:Accession: AE1449			
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.			
D.; Jones, L.M.; Karst, U.			
Science 294, 849-852, 2001			
ok, C.; Schluter, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma			
ok, C.; Schluter, J.; Sincos, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,			
A:Title: Comparative genomics of Listeria species.			
A:Reference number: AB1077; MUID:21537279; PMID:11679669			
A:Accession: AE1449			
A:Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-1946 <GLA>			
A:Cross-references: UNIPROT:Q92FH4; GB:AL592022; PIDN:CAC9365.1; PID:G16412551; GSPDB:C			
A:Experimental source: strain Clip11262			
A:Gene: lin0132			
Query Match			
Best Local Similarity 20.3%; Pred. No. 2.1; Length 1946;			
Matches 217; Conservative 160; Mismatches 392; Indels 302; Gaps 55;			
Qy	32	ILKQTAQSSISGADYAESGSKLIKINETGSPV-----DDTVDTLFSDKRTTPPEKIDN	86
Db	530	IASTLKAKPSMAGIDLSTYSTERGEF---TAVPINLIDTDSITNI---RVVAVLEEDY	581

RESULT 46  
T13734  
groovin gene protein - fruit fly (Drosophila melanogaster) (fragment)

Query Match	3.6%	Score 182;	DB 2;	Length 4151;
Best Local Similarity	20.0%	Pred. No. 6.4;		
Matches 213;	Conservative 170;	Mismatches 372;	Indels 312;	Gaps 54;
QY	88	AKGPREQE-LKAVTENTESEKQITSGOLEQSKESLSINKTVPSTSNWEICDFTTKGNTL	146	
DB	1605	AKDPKLLERVKAIRE-----ELTNLSKPLQSLKALAKDISAERARAGGDADHIT--SEV	1656	
QY	147	VGLSGKSGVEKLSOTDHLVLPSSQAADGTQLIQVASFAFPDPKKTATIAEYTSRAGENG---	202	
DB	1657	DGLA-----DRMSE-----LQRLDDRCGELQSA-----TAVSQFNEQMKSLGIDLN	1699	
QY	203	---EISQLVDGQKEI-----INE-----GEVFNLSVLLKKVTIPTGY-----	235	
DB	1700	DLETEIKLSPGGEIKIVQVIDDVCKIOTKLDRLVGRLEDAERAAVLVDVAGFAADTT	1759	
QY	236	-----KHIGQDAFVNK-NIAEVNLPESLETISDYAFAHALKQLDLP-----	278	
DB	1760	QTRREQISTLRKTGR---LNNRPVDHEDNLHSTLKALEF-YDHQSQTLDLDIQDVSDEFK	1815	
QY	279	NLKAIGELAFDNOITQKLSLPROLMRLAE--RAFKSNHIKTIEFRGNSLKVIGEASFQD	336	
DB	1816	RMKFPVG-----SELDOIRRQEDFRNFRERKVEPLAINVDKVNAGRDIVRS	1862	
QY	337	--NDLSQLMLPDGLEKTEISEAFTGPNCGDHYNNR---VVLWTKSGNKPGLA-----	383	
DB	1863	AGSGVSTATEKOLEKLNDR---WDLKERNERDRRLDVALLOSGFQALAGLAKSLWS	1919	
QY	384	-TENTYVNPDKSLWQESPEIDY-----TKWLEEDFTYQKNSVTGFSNKGLOKVYK-	432	
DB	1920	DTHEMWAN-----QKPPSSDYKVVKAQLEQKFLKQMLLDRQNSMGSLANLGEVANHC	1973	
QY	433	--KNLEIPIKQHNQV--TITEIGNAFNVDVFQNKTIKYDLEBEV-----	474	
DB	1974	EPERASIEKQNLDMKRFDALTDGAEQ-----RELDLEAMEVAKRPHDKISPLE	2024	
QY	475	--LPSTTRTKTGAPAFQSNLKSPEASDDLEIBEKEGAFMNNRI--ETLELKDCLVTIGDAA	530	
DB	2025	LWLDNTERSVMKAMEL-----IPTDEEKIQORIREHRDLHDEILGKPFSDSLADVA	2075	
QY	531	PHINHIY-----AIVLPSEVOIGRSAPRQNGANNLI PMGSKVKTGLGEMAFLSNRLEHLD	586	
DB	2076	AQLMHLVSDBEAVNLGKVRGVTE---RYTG---LVDASDNIGAL--LAESRQGLRHLVL	2127	
QY	587	S-----EQOLTEIPVQAFSDNALKEVLLPASIKTIREAPFKNNHLKQLEV-	632	
DB	2128	SYQDLVAMMESMEALRKFSVPVYA-----EKLEQMDHLELNMENIAGHSNVST	2180	
QY	633	---ASALSHIA-----FNALDDNDGD-----EOF---DNK	656	
DB	2181	VESGAELMKHISNDEAIQLDKLDSLQRRYGDLTNRGDLLKSAQNALPLVQQFHEAHR	2240	
QY	657	VVVKTHNSVALADGEHFI VDPDKLSGTIVDLEKILKLEIGLDYSTLRQTQTQFRDMTT	716	
DB	2241	LVENMQSAEALAPSEPRQADVLRLGEGELADMPILDSINQVG----POLCOLSPGEGAA	2296	

Db 371 KIEATAKTVNFDLGLFTD---ABEEVYLREKAKWAG-----TIIIPESTKSGAT 418  
QY 551 S-----AFRQNGANNLI FMSGKVKTLGEM-----AFLSNRLE 582  
Db 419 PGKTVPTLKETYPHGISVALAENSIVYELIEKIGSDETFGDLQNPDGKQPKKGILINETK 478  
QY 593 HLDLSEQQLTEIPVQAFSDNALKEVLLPASLKTIREAFKPNKHUKLEVASALSIAFNP 642  
Db 479 RKELLE-KIMNKIKIE-----BDKLP-NLKKEYEBKYVYEAKVNEPKPAFNHFEA 528  
QY 643 ALDNDGDEQFQDNKVVVTHNSYALADGEHFI VDPDKLSSTIVDLKILKLEGLDYST 702  
Db 529 RLDNTLVENKDD---FKKREAYMEKKULESCSYEQNSNLINKLKKQLYLE---DVLV 583  
QY 703 LRQTQTQTFRDMTTAGKALLSK-----SNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEB 757  
Db 584 RKDIADDEIKHFSFMWKLSKSEIYDLAQEIRKNENKLTIENTKDFSGVVEL-----QVQK 638  
QY 758 ALVTYK--ATKNGQLLERSINKAVLAYNNSAIKKNVGRLEKELDLLTGLVEGKGPLAAQ 815  
Db 639 VLIIRKIKKALKNVQNL-----LKNKAVK----- 661  
QY 816 TMVQGVYLLK---TPLPLPEYVIGLVNVEDKSGKLIYALD-----MSDTIG 858  
Db 662 ---DLYIIPKVKYTKSEKPEPYVLMVKREIDKLKDFIPKISMIATEKKNKPTVAADIVA 718  
QY 859 EQQK-----DAYGNPI-----LNVDEDNEGYHALAV---ATLADYEGLDIKTILNSKLS 904  
Db 719 KGQSLRGASGTGTTGNTVNAQTAVVQPHQVNVVTVQPGTIGHQAQGEAETQINS--V 776  
QY 905 QLTSLRQVPTAAYHRAGIFQAIONAAAEABQLLPKPGTHSEKSS 948  
Db 777 QAAQVQQTQAGAGQGVASTQTISQAPKT-QASPEPAPAAPPST 819

RESULT 48  
OVBX  
adenylate cyclase (EC 4.6.1.1) - yeast (*Saccharomyces cerevisiae*)  
N/Alternate names: ATP pyrophosphate-lyase; protein J1401; protein YJL005w  
C/Species: *Saccharomyces cerevisiae*  
C/Date: 28-Dec-1987 #sequence revision 08-Sep-1995 #text\_change 09-Jul-2004  
C/Accession: S56776; S56775; A24776; S05828; S55183  
R/Rio Van, D.; Ferreira, J.; Jacq, C.  
submitted to the Protein Sequence Database, September 1995  
A/Reference number: S56776  
A/Accession: S56776  
A/Molecule type: DNA  
A/Residues: 1-1821 <DEH>  
A/Cross-references: UNIPROT:P08678; EMBL:Z49280; GSPDB:GN00010; MIPS:YJL005w  
R/Ride Haan, M.; Smits, L.A.; Grivell, L.A.; submitted to the Protein Sequence Database, September 1995  
A/Reference number: S56771  
A/Accession: S56775  
A/Molecule type: DNA  
A/Residues: 673-2026 <ZAG>  
A/Cross-references: EMBL:M12057; GSPDB:GN00010; MIPS:YJL005w  
R/RKataoka, T.; Broek, D.; Wigler, M.  
Cell 43, 493-505, 1985  
A/Title: DNA sequence and characterization of the *S. cerevisiae* gene encoding adenylate  
A/Reference number: A24776; MUID:86079531; PMID:2934138  
A/Accession: A24776  
A/Molecule type: DNA  
A/Residues: 1-261, 'L', 263-547, 'L', 549-591, 'H', 593-708, 'I', 710-961, 'P', 963-1387, 'S', 1389-  
A/Cross-references: EMBL:M12057; NID:g171359; PIDN:AAA34549.1; PID:g171360  
A/Note: the authors translated the codon TTA for residue 262 as Ser, ACG for residue 311  
2 as Leu, TCA for residue 1566 as Ala, GAG for residue 1659 as Ala, GGT for residue 1738  
R/Mason, P.; Lenz, G.; Jacq, C.; Danchin, A.  
Curr. Genet. 10, 343-352, 1986  
A/Title: Yeast adenylate cyclase catalytic domain is carboxy terminal.  
A/Reference number: S05828; MUID:88165073; PMID:3327602  
A/Accession: S05828  
A/Molecule type: DNA

A/Residues: 1042-1426, 'D', 1428-1460, 'T', 1462-1955, 'V', 1957-2008, 'M', 2014-2015, 'TNFYKWLRT'  
A/Cross-references: EMBL:X03449; NID:G3487; PIDN:CAA27175.1; PID:G3488  
R/Ride Haan, M.; Smits, P.H.M.; Grivell, L.A.  
submitted to the EMBL Data Library, May 1995  
A/Reference number: S55183  
A/Accession: S55183  
A/Molecule type: DNA  
A/Residues: 673-2026 <DEH>  
A/Cross-references: EMBL:X87611; NID:G854567; PIDN:CAA60917.1; PID:G854568  
C/Genetics:  
A/Gene: SGD:CYR1; CDC35; MIPS:YJL005w  
A/Cross-references: SGD:S0003542; MIPS:YJL005w  
A/Map position: 10L  
C/Superfamily: yeast adenylate cyclase; leucine-rich alpha-2-glycoprotein repeat homolog  
C/Keywords: CAMP biosynthesis; duplication; phosphorus-oxygen lyase; tandem repeat  
F/669-1343/Region: leucine-rich 23-residue repeats  
F/1065-1087/Domain: leucine-rich alpha-2-glycoprotein repeat homolog <LRR>  
F/1610-2026/Domain: catalytic #status predicted <CAT>  
F/1664-1749/Domain: yeast adenylate cyclase catalytic domain homolog <YACC>

Query Match 3.6%; Score 181.5; DB 1; Length 2026;  
Best Local Similarity 20.1%; Pred. No. 2.3;  
Matches 223; Conservative 145; Mismatches 365; Indels 375; Gaps 48;

QY 29 KEPILKQTA-----SSSI---SCADYABSSGSKLKINETSGPV----- 65  
Db 616 KTPILNKTYSYTKFTSSVNMNSPDQAQSSG--LLQDEKDDVRCQLEHYKDFSDLD 673  
QY 66 -----DITVTDLPDSKRTTPEKIDKNLAKPREQELKAVTENTESEKQITSGS 113  
Db 674 PKRYAIRFNFTDFTTSLCTPATVVEIIPALKR-----KNVITAAQ 717  
QY 114 QLBSQESLSLNKTVPTSTNWEICDFITKGNVLVGLSKS-----GVEKLSQT----DHL 163  
Db 718 NFQISLVGKLSKILRPTSKPILE--RKLILLNGYRKSDDLHIMGIEDLSFVFKFLPH 775  
QY 164 VLPQOAGDTQLIQVAFPTPDKKTAETSRAGENGESQLDDVDGKEIINEGEVNS 223  
Db 776 VTPSH-----PTPEQQRIM-----RSEFVHVDLRNMDLTPPIIF-- 811  
QY 224 YLLKKVITPTGYKHIGODAFVDNKNIAEVLNPELETISDYAFALHAKQLDL----- 276  
Db 812 -----YQTSIESLDVSNNAIFLP--LEFT-ESSIKLSLRWNIARSKPPS 857  
QY 277 -----PDNLKAIGELAFDQNTGKLSLPR-----QLMRLA 307  
Db 858 NITKAYKLVSELQRFNFKVPNSIMKLSNLTILNLCNELESPLAGFVELKNLQLDLS 917  
QY 308 ERAP-----KSNHIKTI-----EPRGNSLVIGESQD 336  
Db 918 SNKEMHYPEVINYCTNLLQIDLSYNKIQSLPQSTKYLVLKAKMNLSHKLNFI GDLS-EM 976  
QY 337 NDLSQLMLPDGLEKIESEAPTPGDDHNNRVLTGSKGNKPSGLATENTYVNPDKSL- 395  
Db 977 TDLRTNLN-----RYNRISIKTNASNLQNLFLTDNRISNPDTL 1017  
QY 396 -----WQESPEIDYTKWLEEDFTYQKNSVTGFNGK-----LQKVKNK 434  
Db 1018 KLRALIEQENP---ITSISPKDF-YPKNMTSLTLNKAQLSSIPGELLTKLSFLEKLELNQ 1073  
QY 435 N--LEIPKQHNGVITTEIGDNAPFNVDF-----ONKTLRKVDLEEVKLPSTIRKIGAF 486  
Db 1074 NNLTRLQPEISKVLKLVLSVARNKLEYIPPELSQLSLRTLDL----- 1117  
QY 487 FQSNLKSF-EASDDLE-----EIKEGAFMNNRIETLELKDVLVTIGDAAPHNNHYAVL 541  
Db 1118 -HSSNIRDVFGMENLELTSLNISNAFGSSLEN-----SFYHNSYG--- 1160  
QY 542 PESVOEIGRSFRONGANNLI-----FMGSKVTKLGEMAFNLSNRLSHLDLSEKQ 591  
Db 1161 ---SKLSKSLMPTFIADNQPDAMWPLFNCFVNKLVLNLSYNNP--SDVSHMKL---ES 1211  
QY 592 LTEIPVQAFSDNAL-----KEVLLPASIKTIREAPKKNHLKQLEVASALSH----- 638

Db 1212 ITL-----YLSGNKLTLSGDTVLKWSLTKL-----MLNSNQMLSLPAELSLSLSQSVFD 1263  
Qy 639 IAFNALDDNDGDEQD-----NKVVVKTHNSYALADGEHFI---VDPDKLSSTIVDLK 690  
Db 1264 VGANQLKYNISYHNDWNRNKKELKYLNSGRRPEIKSFISHDIDADLSLTVLPLQK 1323  
Qy 691 ILKLJEGLDYSTLRQTTQTFQFRDMMTAGKALLSKNLRQGEKQKFLQEAQFPLGRVDLDK 750  
Db 1324 VLGLMDVTLNTTKVPDENVNFRILTASII-----NGMRVGVADTLGQRDYVSSRDVTFER 1379  
Qy 751 ATAKAEKALVTKATNG-----OLLER----- 773  
Db 1380 FRGNDDCECLLHDSKQNDYGHNISRIVRDIYDKILIRQLERYGDETDNDNIKTALRFS 1439  
Qy 774 --SINKAVLAYNNSAIKANKVRLEKELDLTG-----LVGK-----GPLAQATM 817  
Db 1440 FLQNLKINGMNSVDNGADVANL--SYADLLSGACSTVIYIRGKKLFAANLGDGMALISK 1498  
Qy 818 VQGVYLLKTPPLP-----EY-YIGLNVYFDKSKLIYALDMSDTIG--EGQKDAYGNPIL 870  
Db 1499 NNGDYQTLTKQHLPTKBEYERIRISGGYVNGKLDGVVSRVAVGFPDPLPHIASPDI 1558  
Qy 871 NVDENEGHALAVATLADYEGLDIKTI 898  
Db 1559 SVVTLTKADEMLIVATHKLWEYMDVTV 1586  
RESULT 49  
AGI085  
Hypothetical protein lmo0086 [imported] - Listeria monocytogenes (strain EGD-e)  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AGI085  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AGI085  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1959 <GLA>  
A:Cross-references: UNIPROT:Q8YAN1; GB:NC\_003210; PIDN:CAC98301.1; PID:gl6409445; GSPDB:  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0086  
Query Match 3.6%; Score 180.5; DB 2; Length 1959;  
Best Local Similarity 19.2%; Pred. No. 2.5;  
Matches 206; Conservative 161; Mismatches 384; Indels 321; Gaps 52;  
Qy 20 HQGEVFSLVKBPILKQTOQASSISGADYAESGSKLKINETSQVDDTV--TDLFSDKR 77  
Db 199 NNHYFDLKKELMLVSGTAKETDID-----IRSIKVEIDDAEKRSLKTAKE 244  
Qy 78 TTFEIKONLAKPREQELKAVTENTESEKQI--TSGSQLEQSKS--SLSLNKTVPSTSNW 134  
Db 245 TEKETTKDSTEK-----ESEKQNDASGQKSAQNNESESQDQOTQDAS-- 288  
Qy 135 EICDFITKNTVLGSKSVEKLSQTDHLVPSQAADGTQLIQVASFAPTPDKTAIAEY 194  
Db 289 -----GNNTNGSANNTARSNGTGG---SNSDNGSQ----- 317  
Qy 195 TSRAGENGISQLQVD-GKEII---NEGEVFNFSYLLKKVTIPTYGHIGQDAFVNKNTA 250  
Db 318 --PGGTQKNPNSGVDNKAKEIIKKLENAENTTF-----QVP-----IV 354  
Qy 251 EYNLPESLETISDYAFALHAKQIDLPDLNKAIG--ELAFPD--NQITGKLSPLQMLRL 306

Db 355 DVNL-----TVK-----GQVAAAKLKLTDSSSKRLNSLEAILYDSKNNIVKKEKLN---STK 402  
Qy 307 AERAFKSNHIITIBFRGNLSKLVIGRASFDNDLSQLMLPDGLEKTESEAFTCNPGDDHYN 366  
Db 403 ANQNFTHNLK-----YGETYQVVVGSYK----- 427  
Qy 367 NRVLWMTKSGKNSPGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT-----G 421  
Db 428 -----SASNKNQDITPFRQTV-----BAKPVVLTPKVTGERCEYMAKELTATELYG 473  
Qy 422 FSNKGLQKVKKNKLEIPKQNGVTIIEIGNAFNVPFQNKTLRK---YDLEEVKLPST 478  
Db 474 HIDKLVLKIKENNSVITTSQKVTVDASQTKDQGVKFNLSDDKEYIIIEEELIVDGK 533  
Qy 479 IRKIGAPAFQNNLKSPEASDLB---EIKEGAFMNNRIETLELXDKLVTITGDA-- 530  
Db 534 NVTDDGTFISTLKAKETIEGLNLSYSTDKGEFVSPNLDVDRDESITSRYVAYLEDD 593  
Qy 531 FHIN-----HIYAIV---LPESVQEIQRSAFRQNG-----AN 559  
Db 594 YKVGNSNAKEYAVSVVDANQKTAVKVGRITVDMNDGNIVFGYISGNNQSDYTFATPAS 653  
Qy 560 NLIPIWGSKVTLGEMAFNLSRLEHLDLSEQKQLTPIVQAP--SDNALK--EVLLPASLXTI 617  
Db 654 NSVVVGKTKTPTVEFS-----LKEAQDKLT--INYEVDADNTLLFDNLTHPTLKLY 704  
Qy 618 REEA---PKKXHLKQLEVASALSHTAFNALDDNDGDEQFNKVVVVKTHNSYALADGEHF 674  
Db 705 KSDAQMGYSGNPVATVDTLKXSD---ITNLLFDFGLSEAYYVVVMT--GSYNLDDGAGI 759  
Qy 675 IVDPKLSSTIVDLKELKILJEGLDYSTLRQTTQTFQFRDMMTAGKALLSKSLNR----- 728  
Db 760 MVDELIGQSSVPRTEITIKVNASFSLDSV-DTTKVAINVKLSDAVNLNDANLKIYEKT 818  
Qy 729 -----QOEKOKFLOEA--QPFGLRVDLDKA-IAKAEKALVT--KKAATKNGQLLERS 774  
Db 819 NTLVKVTPLHGDFEKLMSDGSNLSFENLSINKEYLVKVEDGYDSGMNHPVPEGQFIPT 878  
Qy 775 -----INKAVLAYNNSAIKANKVRLEKELDLTLGLVEGKGL-----AQATWQGVYL 823  
Db 879 KKEPETDKVLDDY-----KASEMKLGGLAGIAATEKPLDIDYNAVSSITYSIYK 928  
Qy 824 ---LKTPLPLPEYITGL---NVYFDKSGKLIYALDMSDTIGEGOKDAYGNPILNV-- 873  
Db 929 ADDNLTPFVEQEVSTAADFEXNVFDTNKL-----GRGYSY---VIKADVVW 974  
Qy 874 EDNEGHALAVATLADYEGLDIKTLNLSKLSQTSIROVPTAAHY--RAGIFQAIONAAA 931  
Db 975 NDNYEDHHIEISS---DTIQIK-----KEKPTVEYBILSRITASEIKLVVV 1017  
Qy 932 EAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKGS 983  
Db 1018 EDEESIVPGTLEITSTTG-----GNEQLQSGKNNVTLF--LSSEGT 1057  
RESULT 50  
S02771  
myosin heavy chain A [similarity] - Caenorhabditis elegans  
N:Contains: myosin Arpase (EC 3.6.4.1)  
C:Species: Caenorhabditis elegans  
C:Date: 31-Dec-1993 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C:Accession: T23622; S02771  
R:Harris, B.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: Z19773  
A:Accession: T23622  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1992 <WIL>  
A:Cross-references: UNIPROT:P12844; EMBL:Z78199; PIDN:CAB01576.1; GSPDB:GN000023; CESP:KI  
R:Dibb, N.J.; Maruyama, I.N.; Krause, M.; Karn, J.  
J. Mol. Biol. 205, 603-613, 1989



A>Title: Sequence analysis of the complete *Caenorhabditis elegans* myosin heavy chain gene  
 A/Reference number: 802771; MUID:89178677; PMID:2926820

A/Accession: 802771

A>Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-116,140-1992 <DIB>

A/Cross-references: EMBL:X08067; NID:g6798; PIDN:CAA30856.1; PID:g6799

C/Genetics:

A/Gene: myo-3; CESP:K12F2.1

A/Map position: 5

A/Introns: 46/1; 192/1; 292/1; 468/2; 1921/3

C/Superfamily: myosin heavy chain; myosin motor domain homology

C/Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle coiled coil

F/89-802/Domain: myosin motor domain homology <MMOT>

F/202-209/Region: nucleotide-binding motif A (P-loop)

F/690-712/Region: actin binding #status predicted

F/793-807/Region: actin binding #status predicted

F/875-1992/Domain: coiled coil #status predicted

F/875-1189/Region: S2

F/1190-1992/Region: light meromyosin

F/153/Modified site: N6,N6-trimethyllysine (Lys) #status predicted

F/208/Binding site: ATP (Lys) #status predicted

F/730,740/Active site: Cys #status predicted

Query Match 3.6%; Score 180.5; DB 1; Length 1992;

Best Local Similarity 20.2%; Pred. No. 2.5; Mismatches 354; Indels 237; Gaps 43;

Matches 187; Conservative 149; Mismatches 354; Indels 237; Gaps 43;

QY 2 KKHKTVALTTTTSVVVTHNOE--VFSLVKEPILKQTQASSISGADYAESGSKSLKIN 59

DB 987 KKHVQDLSLRKARQEQSRDHNIRSL-QDEMANQDEAVAKLN--KEKHQEEENRKLN 1043

QY 60 ETSQGVDDVTVDLFSDKRTTPEKIKDNLAQGRQELKAVTENTSEKQITSGSOLESK 119

DB 1044 EDLASEEDKVNL-----EKIRNKL-----EQMDLEENIDREKR--SRGDIEKAK 1088

QY 120 ESLSLNKTVPSTSNWEICDFTKGNLTGLVLSKSGVE-----KLSQTDHLV 164

DB 1089 RKVEGDLKVAQ-----ENIDEITK-----QKHVDVETTLKKRBEIDLHHTNAKLAENNSII 1137

QY 165 -----LPSQAADGTQLIQV-----ASPAFTPKKTAIAEYTSRAGENG--EISQL 207

DB 1138 AKLQRLIKELTARNALAELEAEARNRSQSDRSSEAREELEELTERLEQGGATAQL 1197

QY 208 DVDGKEIINEGEVP-----NSYLKKVTIPTGYKHIGQDAFVNKNIAEVNLPSELET 260

DB 1198 EANKK---REABIAKLREKEBEDSLNHETAISSLKRHGD-----SVAE--LTEQLET 1245

QY 261 ISDYAFALHALK---QIDLPNKLKAIGELAPFDNQITGKLSIPROLMLAERAFKSNHIK 317

DB 1246 LQKLKAKSEAKSKLQRLERSQAT-----DSEVRSQDLEKAL-----K 1286

QY 318 TIEFRGNSLKV-IGBASQDNDLSQLM-----LPDGLKIEBSEA----- 355

DB 1287 TIEVQYSELQTKADQSRLQDFAALKNRLNNSDNRSLSEMDNQLNSLRLKSTLQS 1346

QY 356 ---FTGNGPDHNNRVVLTWYKSGNPGSLGATENTYVNPDKSLMOESPEIDYTKWLEDF 412

DB 1347 QLDERTNRYDEESRQAL-AATAKN---LEHENTILREHLDEEAES-KADLTROIKLN 1401

QY 413 TYQKNSVTGFNGKLGQVKR---NKNLEIPKHNGVITITEIGDNAPRVNDFONKTLRK- 467

DB 1402 AEIQWKAQFQSEGLNKLEETEAAKALQLKVQE-----LTDNGLFAKIASQEKVRPKL 1457

QY 468 -YDLEVKLPSTIRKIGA-FAPQSNLKSFEA-----SDDLPEEIKEGAFMNNRIETL 517

DB 1458 MODLDAQ--SDVEKAAQVAFYKRRQFESIIAEMKKTKDLSSELDAQQRNQLST 1515

QY 518 EL-----KDKLVTTGDAAFHINHYIYVLPESVQEIQRSPQNGANNLIFMGSKVTL 571

DB 1516 DLFAKTAANDLAEYLDSTRRENKSLAQEVKDLTDQLGEGG-----RSV 1559

QY 572 GEMAFSLNRLEHLDSLQEQKQLTEI PVQAFPSDNALKEVLLPASLKTIR-----BEA 621

DB 1560 ABLOKIVRKLEVEKEEELQKALDEARALAEBAE-KVLRQAIEVSQIRSEIEKRIQEKEE 1618

QY 622 F---KXNHLKQLEVASALSIAF-----NALDDNDGDQFONKVVVKTTHNSYALADGEHF 674

DB 1619 PENTERNHQRALSMQATLEAETKQKEALRIKKLESNDINDLEIALDHANRAYADAQKT 1678

QY 675 IVDPKLSSTIVDLEKILKLEGLDYSTLRQTTO-TQFRMTTGTAGKALLSKSNURQGEKQ 733

DB 1679 I-----KCYMETVQELQFQIEBEQOKQIREQFLASE---KENAILQSEKD 1722

QY 734 KFLQEAQFFLGRVLDLKAIAKAE-----KALVTKKATKNGQL-----LE 772

DB 1723 ELAQAE-----AARARNNAEAEIELREONQNDLNAHVSAUTGQRKLEGLLAHAHLE 1778

QY 773 RSINKAVLAYNNSAIKKANVKRLEKEL 799

DB 1779 EIANELKNAVEQGGKASADAARLAEL 1805

RESULT 51

JC6009

surface-located membrane protein Imp3 precursor - Mycoplasma hominis

C/Species: Mycoplasma hominis

C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004

C/Accession: JC6009

R/Ladefoged, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.

J. Bacteriol. 178, 2775-2784, 1996

A/Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene system

A/Reference number: JC6009; MUID:96213016; PMID:8631664

A/Accession: JC6009

A/Molecule type: DNA

A/Residues: 1-1302 <LAD>

A/Cross-references: UNIPROT:Q49547; EMBL:X95601; NID:g1197335; PIDN:CAA64858.1; PID:g119

C/Genetics:

A/Gene: lmp3

A/Genetic code: SGC3

C/Superfamily: surface-located membrane protein lmp3; tetratricopeptide repeat homology

C/Keywords: duplication; membrane protein

F/1-24/Domain: signal sequence #status predicted <SIG>

F/25-1302/Product: surface-located membrane protein lmp3 #status predicted <MAT>

F/957-992/Domain: tetratricopeptide repeat homology <TT1>

F/983-1026/Domain: tetratricopeptide repeat homology <TT2>

F/1089-1120/Domain: tetratricopeptide repeat homology <TT3>

F/1154-1190/Domain: tetratricopeptide repeat homology <TT4>

Query Match 3.5%; Score 180; DB 1; Length 1302;

Best Local Similarity 18.6%; Pred. No. 1.4;

Matches 199; Conservative 183; Mismatches 417; Indels 269; Gaps 46;

QY 40 SSISGADYAESGSKSLKIN--ETSGPVD---DTVTDLFSDKRTTPBKIK-----DNLAK 89

DB 125 NSNDQGVNDSDAKALNENQIDSLPIDIKTKTNENLENAKELLNKAERELQSKIF 184

QY 90 GPREQLKAV--TENTES-----EKQITSGSOLESKESL-----SLNKTV 128

DB 185 NEKKQELKRVLDLEDTKEVDFTKEQKVFETNINETSIEDIKNKILIEVEKATSSLSKI 244

QY 129 PSTSNWEICDFTKGNLTGLVLSKSGVEKLSQTDHLVLPQQAADGTQLTQVASFAPTPDKK 188

DB 245 LNTQQLQELQEF---ENIKKQLQDFINTKLNDAKQYSIKQKALDKINSLINGI-----NKN 295

QY 189 TAIABYTSRAGENGIEISOLD-----VVG-----KEIINGEVFNSVLLKKVTIPT 233

DB 296 STIKEL--KAGQNALIKAKBEAGLEKEKLDQNKIDTLKETINNAKEP-----KKLLIDN 348

QY 234 GYKHIGQDAFVDNK-NIAEYVNL---PESLETISDYAFALHALKOIDLPDLNKAIGELAFF 289

DB 349 DQKIVDLKSLNDNEISAKESLSKDKESMESANDL-----LNTKLIYEKILINKFNQ----- 400

QY 290 DNQITGKLSLPROLMLAERAFKSNHIKTIEFRGNSLKVGEASFQDNDLSQLM--LPDG 347

DB 401 -----EKEAKFNELE--QTRQNIENFLTDEVKNPNYATLVKDLTNA 440

QY 348 LEKISEAFTGNPGDDHYNNRVVLTWTKSGKNPSPGLATE-----NTVNPDKSLW- 396  
Db :  
441 KDAKSVTNSNKSNDIIAANEALIQALADANKAQDVEANKSIKEQLNALIDRANTLLP 500  
QY 397 ---QESPEIDYTWLEEDFTYQKNSVTGFSNGKLQVKRNKNLEIPQHNGVT--ITEIG 451  
Db :  
501 QLNDDNSEIVKAKESLNAEITNANKAVQNNDNASMQSAKSLDDKVTKIQHQTETFNCK 560  
QY 452 DNAFNRVDPNKTLRKYLDEEVKLPSTIRKIGAFAFOSNN-----LKSF-EASDDLBEI 504  
Db :  
561 DAKFKELEQTRKDIDNFLTDGVK-----NNPNYATLVKDLTNAXDDKKSV 605  
QY 505 KEGAFMNNRIETLELKOKLVITGDGAAPHINHHVAIVLPESVOEIGRSAPRO-----NGAN 559  
Db :  
606 TKS-----SNKSEITAANDLEKQALDK-----AKVAKQIDEANKSIKEQLSDSITVAN 654  
QY 560 NLIIFMGSKVKTILGEMAFNLNRLEHLDLSEKQLTEIPQAQPSDNALKVELLPAKLTIRE 619  
Db :  
655 QLJ-----NKLVDSKDIOAKTELSEIQSASEQLNINNPTSMQSACE 698  
QY 620 --HAFKQNHCLKOLEVASALSASHIAFNALDDNDGD--EQFDNKVVVTHNSYAL-----A 669  
Db :  
699 SLDAKVTEITTKLETNFNDKDVKFKELETRKDIDEFIN--TNKTNPNYSTLISELTSKR 756  
QY 670 DGEHFIVDPDKLSSTIVLDLEKI-LKUIEGL-----DYSTLRQTTOTOFDRMTTAGKA 720  
Db :  
757 DSKNSTVN----SSNKSDETANTELKQALAKANTDKAADNLAKSTKEQLNNGISSANT 812  
QY 721 LLSKSNLRQGEKOKFLQEAQFFLGRVDLDKAIKAEKALVTK-----KATKNGQLLE 772  
Db :  
813 LLAKUT---DKDNTIQQA-----KTELEKEVKANOAVASNNTASMOSAKSLDAKYTE 863  
QY 773 RSINKAVLAYN-NSAIIKANVRKLEKELD-----LLTGLVEKGKPLAQ 814  
Db :  
864 --ITKLETNFNDKDVKFKELEQTRKDIDEFINTKTNPDYSTLISELTSKRDSKNSITN 921  
QY 815 ATMVGCVLLHTPLPUPEYYIAGNVDFPKSGKIYALDMSDTTGEGEKDAYGNPINL--- 871  
Db :  
922 SSNKSIDIETANTELK-----QALAKANTDKQADNLARSTKEQLNKSISANT 969  
QY 872 ----VDENEGYHA-----LAVATLADYEGLDIKTILNSKLSOLTSTIRQVPT 914  
Db :  
970 LLAKUTDKDNTIQQAETEKEVKANOAVASNNTASMOSAKSLDAKYTEIT--KKLET 1027  
QY 915 AAYHRAGIFAQIONAAAAEQLLPKPGTH-----SEKSSSSSESANS 955  
Db :  
1028 FNKDKVFKRELEQTRKDIDEFINTKTNPNYSTLISELTSKRDSKNS 1075

RESULT 52  
A90551  
C:Species: Mycoplasma pulmonis - Mycoplasma pulmonis (strain UAB CT)  
C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C:Accession: A90551  
R:Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001  
A>Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
A:Reference number: A95512; MUID:21267165; PMID:11353084  
A:Accession: A90551  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2819 <CUR>  
A:Cross-references: UNIPROT:Q98QP8; GB:AL445566; PID:g14089727; PIDN:CAC13486.1; GSPDB:B-C  
A:Experimental source: strain UAB CTRP  
C:Genetics:  
A:Gene: MYPU 3130  
A:Genetic code: SGC3

Query Match 3.5%; Score 178.5; DB 2; Length 2819;  
Best Local Similarity 20.0%; Pred. No. 5.2;  
Matches 211; Conservative 163; Mismatches 387; Indels 295; Gaps 52;

Qy	2	KXHLKTVALTUTTVS-----VVTHNQEVSLVKEPILKQTOQASSISTGADYAESSGKSKL	56
Db	611	KSELKTTLSIKFTGAFIONKILSHN-----LSELVLR-----IVSG-----	646
Qy	57	KINETSFPVDDTVTDLFSDKRT--TPPKIKDNLAKGREQELKXAVNTSEKQITSGSQ	114
Db	647	--EKTFAIIEKLIIDDFIDFNKAVYQKVEKIEDFUFVFNTE--KAITDFVR-----	693
Qy	115	LEQSKESLSLNKTVPSNSNWEICDFITKGNTLVGLSKSGVEKLSQTHLVLPQQAADGTQ	174
Db	694	-----FVLNELVAKENPK-----ALVSFGISQFLNSNNEVISTQ-----	728
Qy	175	LIQVAFAPDPDKTAJAEYTSRAGENGESQLDVGDEKEIINGEVENSYLLKKVTIPTG	234
Db	729	-----TSQAFYEVSKTLV-----NIVDTEE-----LLEKV-LKNF	757
Qy	235	YKHIGQDAFVNDKNIAEVLNPESLE--TISDYAFAH-----LALKQIDLPNLKAIGELAF	288
Db	758	IEEIKSSAQLKLSNPKFENLIEKLOKSIIISFTGTDLGWSFLKQVVKSKLSVSEQRDK	817
Qy	289	FDNOITCKLSLPRQLMLAERAFKSNHIKTIEFRGNSLKVIGAS-----FODND--	338
Db	818	FTSLSKEVVSF-----VFKSENKN-----KUISAGTYLSNIEFLKENSVDV	859
Qy	339	-----LSQLMLPDGLEKIESEAFTEGNGDDHNNR--VVLTWTKSGKNPGLATENT	387
Db	860	FKGIFSHILNFEKTPPELLNKIIDVINSWESIIOYNDAPFVLNFKT--NSSWF--KT	914
Qy	388	YVNPDKSLWQB--SPEIDYTKWLEEDPTYQKNSVTGF--SNKGLOKVR--NKNLEIPKQ	441
Db	915	YV--KDISFVSNSEIOYANILTSKLN-KLASQAGFELTNSVSEGIKNTFTNLTKLIAD	970
Qy	442	HNGVTITEIGDGNFRNVDFONKTLRKVDLEVKLPSTIRKIGAFQSNLKS--FEASDD	500
Db	971	SN--TIDALVD-TPVN-----SKTL--LEIDKKNPFIASLKNOIQIFALSDDDFFSVFKALIS	1021
Qy	501	LEEIKEGAFWN-----NRIETLEKDKLVITGDAAFHINHIIYAVILPESVQ-----	546
Db	1022	SNKLKDAFVNOYKQDITNVLKELLPEKF-----FNLIYSFMPMKTSELFENNPE	1072
Qy	547	-----BIGSAFRONGANNLIIPWGSKVITLGEMAFLSNRLEHLDISEQKOLTEIPVOAF	600
Db	1073	ALNKTTTTILRNALKNLDPOFI--NKQMDI-----IPDNLDKYQKLENYQVVLGLILKON	1125
Qy	601	SDNALKEVLPASIKTIREAFKKNHLKQLEVASALSASHIAFNALDNDGDGEQDNKVVVK	660
Db	1126	SDAIATKV--QAIPIKQVNEANFODVLKVLVLSQTKKFMGFELNADETNIKISDTLTKILK	1183
Qy	661	-----THNSYALADGEHIVDPDKLSITIVLEKILK--LIEGLD-----YSTLRQT	706
Db	1184	QVFESNPFKSLALATKDSLVIDGVQANKKIPDTPFLNTFLLSEKDEKTKENYQMVXSI	1243
Qy	707	TOTOFRDMTTFAGKAL-----LSKSNLRQEQEKQFLQEA--QFFLGRVLDLKAIAKPK--	757
Db	1244	LNTKLIDISQNSSESIKNLLTFTKSALSRDE--FYEKTIKVLSEKELASATVPVVEFV	1300
Qy	758	--ALVTKATNGOLLERSINKAVLAY-----NNSALKANVVRLEKELD	800
Db	1301	KFVLKNEKTTNPSDLLDSLPKDFSQOQANSFEELIYLVLRNNKTTVNNYLKLSFEIK	1360
Qy	801	LLTGLVEGKPLAQATWQVGVLLTKPLPLPEYVIGLVYFDKSGKLIYA--LDMSDTI	857
Db	1361	-----SNIIORLLGQITINALVTP-----NGEKIFSDTELOQSITL	1396
Qy	858	GEGOKOAYG--NPILNVDE-----DNEGYPHALAVATLADYEGLDIKTILNSKLSQLT	907
Db	1397	INESLDLFGKNTIISNLYDLVLVSALISINQSSSVLTQVNIITHKALDQO--NSYPAVVK	1453
Qy	908	SIRQVPTAATHRAGIFQAI-----QNAAAAEQLLPK	939
Db	1454	TLISVREGDVKDEFAKAIKKFNLNFAKNDYLLTK	1489

RESULT 53

G41662

130K surface exclusion protein Sec10 precursor - Enterococcus faecalis plasmid pCF10

C/Species: Enterococcus faecalis

C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004

C/Accession: G41662

R/Kao, S.M.; Olmsted, S.B.; Vikenins, A.S.; Gallo, J.C.; Dunny, G.M.

J. Bacteriol. 173, 7650-7664, 1991

A/Title: Molecular and genetic analysis of a region of plasmid pCF10 containing positive terococcus faecalis.

A/Reference number: A41662; MUID:92041679; PMID:1938961

A/Accession: G41662

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-891 <KAO>

A/Cross-references: UNIPROT:Q04111; GB:M64978; NID:g150552; PIDN:AAA65847.1; PID:g150554

C/Genetics:

A/Genome: plasmid

Query Match 3.5%; Score 178; DB 2; Length 891;

Best Local Similarity 19.0%; Pred. No. 0.99;

Matches 201; Conservative 176; Mismatches 411; Indels 270; Gaps 47;

Qy 7 TVALTLTVSVVTHNQEVPSLVKPEILKQTCASSISGADYAESGSKLKNETSQVDP 66

Db 45 TVKATQTEQAITEKQOQVT-EKQAVDQKQ-----QVADTAKKEK----- 84

Qy 67 DTVTDLFSDKRTTPEKIKDNLAKPREQELKAVTEN---TESEKQI--TSGSOLEQSK 120

Db 85 DAIDQSVKQQAQVVDQNKDAL-----DQQAQVTDQAQVDEAKKVDEATPSAIEKAKE 139

Qy 121 SLSLNTKVPSTSNWEICDFTKGNLVLGSKSGV-EKLSQTDHLVLPQAAQTQLIQVA 179

Db 140 QVATD-TQAQVDEQKQVVD---QAQTDVNOQAQVVDKAKETN-----AA 179

Qy 180 SPATFPDKTAIAEYTSRA-----GENGEISQLDDVOCKEIINEGEVENSYLLKKVITPTG 234

Db 180 KVQNEKQQAQVTAQKQQAQKLEELAKVAEAKVKAEKEQAQKAEALANKQ--KEEAKAKD 237

Qy 235 YXHGIGQDAFVDNKNIABVNLPESLETISDYAFALHALKQIDLPDNLKAIGELAFFDNOIT 294

Db 238 QKTKDDQQAQVADQQTV---VTTSEKQVTD-AKADTAQAQADLTAK-----ENALK 282

Qy 295 GKLSLPQLMRLAERAFKSNHIKTI EPRGNSLKVIGEASFQDNDLSQMLPDGLEKTESE 354

Db 283 DKQAATKQ----AQNTLDNSKEELKHGKHGINKLPKFSADY-DTKLSABEAT-LEKTALE 336

Qy 355 AFTGNP-GDDHYNRVVLWTKSGKNPSGLATENTVVPDKSLMQESPEIDYTKWLBEDFT 413

Db 337 MNKNFPTSKDEKQKQVWV-----DQHQ-----LSAD-- 363

Qy 414 YQKNSVTGPNKGLOKVRKNKNLEIKQHNQGVITITEIGDNAFRVDPQNKTLRKDYLEEV 473

Db 364 -QKKELSVYVITELLNDVRKKL-----GLSLSVSDQSI-----KFAWDIA 402

Qy 474 KLPSI-----IRKIGAFAPQSNLKFPEASDDLEEKEGAF--MNNRIETLEKDK 522

Db 403 KYSDTGEMHDVTAINK-----AAKENGFKQYPGNMYYENLGGGYETENGKSVKTYLQES 458

Qy 523 L-----VTIGDAAPHNIHIAIVLPESVOBEGISAFQNGANNLIPMGSKVKTGLGMAF 576

Db 459 IRKMLVNMFLDQDGLRGVSHLSLL-----QDQKTALE-----GVSLSGENKNSIPKIH 506

Qy 577 LSNRLHLDLSEQKQLTEIPVQAFSDNALKEVLLPASLKTIRREAFKXNHLKQLEVASAL 636

Db 507 ISYGEKLEDSQYQNGEV-----ASWKS-----KEELQQ-EIASNQ 542

Qy 637 SHTAFNALDNDGDG-QFDNKVVVTHNSYALADGEHPFVDPDKLSTTIVDLKILKLI 695

Db 543 EKLTAAQASDAQAASASQAQALNTAKTQTATKELSLV-----HKATLANIQAQV---- 593

Qy 696 EGLDYSLTRQTT--QTFQFRDMTWTAGKALLSKNSLRQGEKQKFLQEAQFFLGRVD--LDKA 751

Db 854 EAVSVQRNQETVWKLHKEQFN---MLSSDLEKLRNLADMEAKFREKDEREQLIK 910  
QY PKOHNGVYITEI---GDNAFRNVDFQNK-TLRKYDLEEVKLPSTIRKIGAFQFQSNL 493  
Db 911 AKEKLENDIAEIKMGSDNSQUTKWNDELRLKERDVEELQKLT-----KANENA 961  
QY 494 SP-EASDDL---EPIKEGAFMNNRIETLEKOKLVITIGDAAFINHIYAIYVLPESVOE 547  
Db 962 SFLQKSIEDMTVKAEOQQAEEAKKHEEKELERKLSL-----EKKME 1005  
QY 548 IGRSAFRONGANNLIIFWGSVKYTLGEMAFLSNRLHLDLSEQKQLTEIPVQAFSDNALKE 607  
Db 1006 TSHNOQOE-----LKARYERATSETKTKEEILONLQKLLDTE-----1044  
QY 608 VLLPASLKTIRBEAFKKNHLKOLEVASALSALSHAFNALDDNDGDEQFNDKVVVKTTHNSYA 667  
Db 1045 ----DKLKGAREE---NSGLLOLEELRKQADKAKAQAOTAEADAMQINQMTKEKTE-----T 1094  
QY 668 LADGEHFI VDPDKLSSTIVDL-EKILKJIEGLDYS-----TLRQTTQTQFR-DMTTAGKAL 721  
Db 1095 LASLEDTKQTNKALQELDTLKENNLKNVBEELNKSKELLTVENQKMEEFKEIETLKQAA 1154  
QY 722 LSKS---NLROQEKQKFLQEAQFFLGRVLDKRAIAKAEKALYTKATKNGQLLE-----772  
Db 1155 AOKSOOLSALQENVKLAEE---LGR---SRDEVTSKQLEERSVFLNNQLEMKKRES 1207  
QY 773 -----RSINKAVLAYNNSAI---KCANVRLKLEKELDLTLGLVEGKGPLAQATWVQ 819  
Db 1208 KEIKDADEKASLOKSISI---TSALTTEKDAELEKURNEVTVLRG---ENASAKLSHVVQ 1263  
QY 820 GYVLLKTPPLPEYIYGLNVYFDK-----SGKLIYALDMSDTIGEGQKDAYGNPLNV 873  
Db 1264 TLBSDKVLEKLVNLEQLKENKROLSSSGNTDTQADEDERAQESQIDFLNSVVDLQ 1323  
QY 874 EDNE 877  
Db 1324 RKNQ 1327

RESULT 55  
H71611  
probable secreted protein PF80565w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: H71611  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Pertes, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: H71611  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1817 <GAR>  
A:Cross-references: UNIPROT:096205; GB:AE001403; GB:AE001362; NID:g3845216; PIDN:AAC7190  
A:Experimental source: clone 3D7  
A:Genetics:  
A:Gene: PF80565w

Query Match 3.5%; Score 177.5; DB 2; Length 1817;  
Best Local Similarity 18.7%; Pred. No. 3;  
Matches 199; Conservative 159; Mismatches 394; Indels 315; Gaps 48;

QY 16 SVYTHNQEVSLVKEPILKQTAQSS---ISGADYAESG-----KSKLKINE 60  
Db 577 NYEQKKQTLIYNDKIKKKNKNSYEMNNNTYMHGYTDTIENRLNKKKRLNVRG 636  
QY 61 TSGPVDVDTDLFSKRTTPEKIKONLAGPREQ---ELKAVTENTSEKQITSGSQL- 115  
Db 637 RTNTLDDIIV---SDHNSYDKY---NTSKHNRKKNHINEMKKQKNNKQNTLFDVQKOME 691  
QY 116 ----EQSKESLSLNKTVFSTSNWEICDFTTKGNTLVGLSKSGVEKLSQTDHLVLPQAAD 171

Db 692 GIGKEKEKKNMNNN-----IFYNNSYNINNNSSYNINNDIYSDNMTSYN 739  
QY 172 GTQLIQ-VASFAFTPDKTAIAEYTSRAGENCEISQOLDVDGKEIINEGVEFNSYLLKKYT 230  
Db 740 NTKYSGVPSYAHVLINKQVNEYQGLPNYN---NMMIKGSHIINELPK-NNYIYEN-- 792  
QY 231 IPTGVKHHGQDAFVNQKIAEAVNLPESETISDYAFALHAKOIDLDPNKLKAIGELAPFD 290  
Db 793 ----NYIGQNVLTNP-LYNKETKDI FYTIIKYLF-----KLIYSPSLKK---RMEFID 838  
QY 291 NOITOKLSLPROLMRLAERAFKSNHIKTIEFRGNSLKVIGEASFOQNDLSQLMLPDGLEK 350  
Db 839 NCMTKTKIFVIRKVCNFKRPFSSNK-KNNKMNDRSDSYVDNISYYDDD-----885  
QY 351 IBESEFTGNPGDDHYNNRVLTWTKSGKNPSGL-----ATENTYVNPDK 393  
Db 886 -----NNNNNNINILKCKKKKREVGGGIRLNGVNDKTRHDDTIDEKYNKNN 935  
QY 394 SLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKNLEIPKQHNGVTITEIGN 453  
Db 936 YLFMNGVDVLYNK---DQLGYKNS-----LDDN 961  
QY 454 AFRNVDFONKTLRYDYDLBEVKLPSTIRKIGAFQFQSNLKSFEASDDLEIEKEGAFMNNR 513  
Db 962 NNNNNYNNNDNIRRHVSVC-----SYRRAHNK-----YDIKEGGSNDIY 1003  
QY 514 IETLEKDKLVITIGDAAFINHIY-----AIVLPESVOETGRSAFRONGANN 560  
Db 1004 TSNIKRNKKTKNTBEI FNIINMLNKEALKNYTVYDVTILYDEFSKLLKGIFEK---NK 1060  
QY 561 LIFMGSVKTKTGEMAFLSNRLHLDLSBOKO-----LTEIPVQAQSD- 602  
Db 1061 CLF-----KLKENYWSQNSYLLHDKIKKCTCLNYQRLLFHEVINLFVYVYKFCNW 1114  
QY 603 NALK---EVLPLPASLKTIRBEAFKKNHLKOLEVASALSALSHAFNALDDNDGDFQFNDKVV 659  
Db 1115 DVLKNYFDILINGS-----EEA-----IKVLEHFR-----NINKEQID--VIR 1151  
QY 660 KTHNSYA-LADGEHFI VDP-----DKLS-----STIVDLKILKLEGLDYSTUR 704  
Db 1152 KSYNNMYEYLSKSKYEHIIDDIINDYNNKINNMERKINIRIIDIIDIFK-----EYLLLI 1206  
QY 705 QT---TQTFQFDMTTAGKALLSKNLQGEKQKFLQEAQFFLGRVLDLQAKAKAKALVT 761  
Db 1207 QOEHTKEGLKKNHIYKSKILFKNPLPSNLLKI-----LCKKKEKIKENLNTN 1257  
QY 762 KKATKNGQL-----LERSINKA-----VLAYNNSAIKKAN 791  
Db 1258 CFSTVNNMLRNDMIKGSFYFSKYSYCFERLLDYAFSLTITFENINFIYIGDVLKLYE 1317  
QY 792 VKRLEKELDILTLGLVEGK--GPLAQATWQGVYLLKTPLEPYIYGLNVYFDKSGKLIY 849  
Db 1318 VD-FKNSLYLLVITIKLHKFINNLFETIKVREI--LTKVSI-----LKNKYTERI-KL 1369  
QY 850 ALDMSDTIGEGQDAYGNPILNVDEDSYHALAVATLADYEGLDIKTLNLSKLSQLSI 909  
Db 1370 LLNCIPLI-----YLDPHNVILINESY-----EYKMDDDDKIIFSKLSPFSLV 1413  
QY 910 RQVPTAAAYHRAGIFQAIQNAIAAEAEQLPKPTGTHSEKSSSESANSK 956  
Db 1414 SKV---VNHK---LRSVYTHYDITDNLNENBEPIHKNTKSKSMNDDTK 1454

RESULT 56  
T28676  
rhoptry protein - Plasmodium yoelii (fragment)  
C:Species: Plasmodium yoelii  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T28676; A45521  
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.  
Mol. Biochem. Parasitol. 76, 329-332, 1996  
A:Title: Comparison of two members of a multigene family coding for high-molecular mass



Db 813 SKQSAFL--ENTQYTNLEEVREYTSFSEFALMKNLKQKGMVQNLLEAEIKOILTDSSVSK 870  
QY 422 FGNKG-----LOKVRKNKLEIPKQHNGVITTEIGD--NAPRVNDFQNTKRLKYDLEE 472  
Db 871 SDDKATPAVKLIQAFESKRPEPESENAQITDDEADQFVSVNVQIRNLRLG--LDO 929  
QY 473 VKLPSTIRKIGAPAFQSNLKSFEASDDLEETKEGAFMNRITLLELKDGLVTIGDAAPH 532  
Db 930 LLL--NARKAG--IQFNQNDRTS-----TNQRLLELV-----EFASH 965  
QY 533 INHIYAIVLPESVQELGRSAFR-----QNGANNLIFMGSKVKTIGE-----M 574  
Db 966 QOHINVLEADTIESKVFALXHYSELQHKHDLLELLCDSLKLRNDNISVENTELNKKL 1025  
QY 575 AFLSNRLEHLDL---SEQOLTEIPQAFSDNALKEVLLPASIKTIRREAFKKNLQLE 631  
Db 1026 NYC/SURIDELETOLENLQNLTSFL-----STMEEQLV--ALQDESERAMVEH-----E 1073  
QY 632 VASALSHI--AFNALDD---NDGDEQFQDNKVVVKTTHNSYALADGEHFIVDPDKLSSTIV 686  
Db 1074 LTLMSFEGEAVVRLDDCLLRSG-----TSGAHTGLDMTKRISGSV 1114  
QY 687 DLKILKLEGL-----DYSILRQTTQTFR-----DMT- 715  
Db 1115 DV--AVNVIEDLKEKLEAAVYKHESNKEYELKOSFNTLFKEKNEFTASSMOKVYADLTK 1172  
QY 716 ---TAGKALLKSNL-----ROGEKQFLOEAOFFLG-RVDLDKAIKAKALVT 761  
Db 1173 LITSCGSABMSLEVENAVFPDFGDSFEMLEAVRKLSELELQSVIDLQSDLS-S 1231  
QY 762 KKATNGQLLERSINKAVILAYNNSAIKKANVRKLEKELDLTLGLVEGKGLAQATMVQGV 821  
Db 1232 SKSNDMEEMTQSLDSTSL-----RELVEKVEGLELESV-----1268  
QY 822 YLLKTPLPPEYVIGLVNFDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGTHA 881  
Db 1269 --FESSQVFLVSLVQ-----KFITEELANLRLK-QLEAKGNELMEI-EBSLLHHK 1319  
QY 882 LAVA-----TLADYEGLDIKTLNLSKLSQLSIRQ-----VPTAYHRAQIF--Q 924  
Db 1320 TKIAGRLBSLQABESLVAVRSELQKSNELSQSEORLLSTREKUSIAVTGKGLIVORD 1379  
QY 925 AIQNAAAAEQLLPKPGTHSEKSSSESANSKDRGL 960  
Db 1380 NVKQSLAEASAKLOK-----CSEELNSKDARL 1406  
  
RESULT 58  
AC1583  
internalin protein (LPXTG motif) homolog lin1204 [imported] - Listeria innocua (strain C  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AC1583  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, P.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1583  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1237 <GLA>  
A:Cross-references: UNIPROT:Q92CG8; GB:AL592022; PIDN:CAC96435.1; PID:g16413678; GSPDB:C  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lin1204  
  
Query Match 3.5%; Score 176.5; DB 2; Length 1237;  
Best Local Similarity 20.1%; Pred. No. 1.9;  
Matches 236; Conservative 179; Mismatches 365; Indels 353; Gaps 64;

QY 1 MKKHLKTVAL-----TLTTSVVVTHNQEVFSLVKEPIKQTOQASSSISGADYAESGK 53  
Db 1 MKRKSSIIIVGIMLFQSLTWPFYITEAKE--NEQKEEINKPSKITKGLT-----48  
QY 54 SKLKINETSQVDDTVTDLPDSKRTTPPKIKDNILAKGPREBELKAVTENTSEKQITSGS 113  
Db 49 NSLYTKTILETGTIDYDSVFPDSALA--KVVAKEATG-----SENT--TQLVTQA 94  
QY 114 QLEQSKSE-----SLSLNKTPVSTSNWEICDF--ITKGNLTVLGSKSGVEKL 157  
Db 95 DLNKIKSLNGYNGKISVLVTGIDLLVNVTSISLNNQVTDISPIDQLPNLVSLS-----VKN 150  
QY 158 SQTDLHLVPSQ-----AADGTOLIQVASPAFTDKK--TAIAEY 194  
Db 151 NQISLLILNAQNOLPKLTTIDIENNPLDNTDIDQPOLVDVKTSGYGLRKLTTVIK- 209  
QY 195 TSRAGENEISQLDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVNKNIAEVL 254  
Db 210 -----NNPELVNLQO---YTIIRNV-----YFSQVASLTKVEL--VNL 241  
QY 255 PE-----SLE-----TISDYAFAPHALKQIDLPDNLKAIGELAFDFDNOITCKLSLPRQL 303  
Db 242 PKVRKNLERNISINELKVTDLAIEDLPLEENELDTV-----FDN-----IQNLP---286  
QY 304 MRLAERAFKSNHIKTIEFRGNSLVIGEASPDNDLSQMLPDLGLEKTESAFTGNPGDD 363  
Db 287 -----NLKTLDSLKNQLEEVLDKTDVENLPLMTLIDIQONLAKILI--NVQDQ 333  
QY 364 HYNRVVLTWTKSGKNPSGLATENTVYVNP--KSLWQESPEIDYTKWLEBDFYQKNSVTGF 422  
Db 334 --POLVDVKTSDYKELSALTTVIAKNPELVNLQO-----YT--IRNVYFSQVASLTKV 383  
QY 423 SNKGLOKVKRKNLE-----IPKOHNGVTITEIGDANFRV-----DFQN 462  
Db 384 ELVNLPKV-RKVNLERNSINELKVTDLAIEDLPLEEN-----ELDTVFDNIQNLPLKLT 437  
QY 463 KTLRKYLEEVEKLPST---IRKIGAFQSN---NLKSFASDDLEIEKEGAFNV-NRI 514  
Db 438 LDLSKNQLLEVVLDKTDVENLPLMTLIDIQONLAKILINVDQPOLVDVKTSDYKELSAL 497  
QY 515 ETELEKO--KLVTIG-----DAAFHINHIYAIVLPE-SVQEIGSAFRQNGANNLI 562  
Db 498 TTVIAKNPELVNLGYPIQMNVYFSQVASLTKVELANLPKRAVRLERNISINQIELNVL 557  
QY 563 -----FMGSKVKTIGEMAFISNRLEHLDLSEQ--KQLTEIPQAFSD 602  
Db 558 SVKDVNLNTNKITNDSIEKFRGMPIT--LATNLNKNQITNINMLDGPFPQMTLTNIDL---612  
QY 603 NALKEVLLPASIKTIRREAFKKNHLKQ--LEVASALSHTAFNALDDNDGDEQFDNKKVVK 661  
Db 613 NSIS--VLPSNLKTKWPKLSISALNQVTLGKAI-----VVDSD-----LIINN 656  
QY 662 HHNSVALADGHEFIVDDPKLSS-----TIVDLEKILKLE-----GLDYS 701  
Db 657 EISNF-----EKLTEPSPISNFGTYANEKITWSSERIKNLTEVSQFSELINVTGIDGT 710  
QY 702 TLRQTTQTFQDMTMTAGKALLSKSNLROGEKQ---KFLQEAQFFLGRVLDLKAIAKAKA 758  
Db 711 FSGKVQO--PFKRSTTTPVFNADSEIHYPQGTETKEAEFLKDVQ---AQTDDLAIKSDFET 766  
QY 759 LVTKK-----ATKNGQLLE--RSINKAVLAYNNSAIKKANVK-RLEKEL-----DLL 802  
Db 767 MVNLKVKGYTVTNLNENLDGKATPKREVIIVIID-AVOGANITVKYDESGNKLAEISIL 825  
QY 803 TGLV-----EGKGLAQATWVQGVYLLKTLPLPEYVIGLVN 839  
Db 826 TGNVEEYSSSAKEILGYTLTTEMPSNAQGEFSLEBQT-VTVIY-SKNPVPKADITV---Q 880  
QY 840 YFDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDEDNEGTHALAVATLADYEGLDIKIL 899  
Db 881 YTTDEG---MELAPTEITLSG-----NVDEN-----YVTTAKTFTGYELIETPSNA 922



Qy 900 NSKLSQTSIRQVPTAAVHRAGIFQAIQNAARAEQQLPKPGT 942  
Db 923 EGFSE-----NAQTFTV-----VYRAI-----KAEPILAKEVT 951

RESULT 59  
A72287  
hypochemical protein TM1182 - Thermotoga maritima (strain MSB8)  
C/Species: Thermotoga maritima  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C/Accession: A72287  
R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.C.M.

Nature 399, 323-329, 1999  
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing  
A/Reference number: A72200; MUID:99287316; PMID:10360571  
A/Accession: A72287  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1170 <ARN>  
A/Cross-references: UNIPROT:O9XOR4; GB:AE001774; GB:AE000512; NID:g4981717; PIDN:AAD3625  
A/Experimental source: strain MSB8  
C/Genetics:  
A/Gene: TM1182  
C/Superfamily: chromosome segregation protein SMC1

Query Match 3.5%; Score 176; DB 2; Length 1170;  
Best Local Similarity 19.1%; Pred. No. 1.8;  
Matches 204; Conservative 180; Mismatches 390; Indels 292; Gaps 52;

Qy 32 ILKQTOASSISGADYAESGSKKLINETSQVDDTVDLPSDKRTTPEKIKONLAKGP 91  
Db 152 LLLSEAAAGVSI-----YREKKKETELNERTKANLDRVXDLFERER-----QMKSLYLKAK 203

Qy 92 REQELKAVTENSEKQITSGSLOESKESLSL-----NKTVPSTN--WEICDFITKGN 144  
Db 204 RAERFKEYTAQELQKLYYGNVLRERKKLFQVEEKKTEKIKNQLKSLVETKWS 263

Qy 145 TL-----VGLSGVSKLSQTHLVLPQSAADGTQIQVASF--AFTPDKTKTAIAEYTSRAG 199  
Db 264 TLRSFQMDQEIERYTK-----LLEDYKQRQNDLVEMKGFYSSKLADSENKYVELSTRL- 318

Qy 200 ENGISQLVDGKEILNEGEVFNYSLLKKVTIPGYKHIGQDAFVDNKNIAEVLNPSLE 259  
Db 319 --DELEKRRERYKXRLSEME---YIPKGVN--GDYERKAKEL-----EKFE 357

Qy 260 TISDYAPAHALKQIDLPNDLKAIGELAPFQNOITKGLSLPQLMRLAE-----RAP 311  
Db 358 KENENLSRPNDEKEF---LRVRDEISKLKQI---LKJENELLRIGETLEDLEKRRKI 411

Qy 312 KSNHKT-----IEFRGNSLKVIG----EASPDNDLSQM--LPDGLFKIESEFTGNPG 361  
Db 412 TENQILTRRRELEDKNEFKEISRRVVELEDBEKKLTTELNAVRELERLSEIGEIRRVNLE 471

Qy 362 DDHYNRVVLTKSGKNPSGLATENTYVNP-----KSLWQESPE-----IDYTK 406  
Db 472 IDAKEKRL-----REIQFEKEMIERDMREYGRFSRAVRVFEKERFPGLVDVVS 521

Qy 407 WL---EEDFTYQKNSVTGFSNKGKQKVRKNK-----LEIPKQHNGVTIT---EIGDNA 454  
Db 522 NLIEVDEKYSIAVSVLGGTAQNTI---VVRNVDTKAIVIEFLKQNEAGRVITLPDLIDGS 579

Qy 455 FRNVD-FQN-KTLRKYLDEYKLPSTIRKIGAFQSNLNKLSFEASDDLEBEIKGAFMNN 512  
Db 580 FNRISGLENERGFVGYAVDLVKFSDLEVLGGFLF--GNSVVETLDDAIRMKKKYRLNT 637

Qy 513 RIETLELXKDLVTTGDAAFHNHIIYAVLPSVQIEGRSAP---RQNGANLIIPWGSVKV 569  
Db 638 RIATLD--GELIS-----GRGAIYTGREERSSNVFERRIKLK 672

Qy 570 TL-----GEMAFLSNRLBHLDLSE---QKQLTEIPVQAFS--DNALKVELL 610

Db 673 HLEQMEETERQIAEKDDDELASLKTQENLKNQBALVQRELFELSRKSSTTKTVLSEIL- 731  
Qy 611 PASLKTIRREAFKKNHLKQLEVASALSIAFNALDD---NDGDEQFQKVVVTKTHNSYA 667  
Db 732 -RSINQLQEEV---ENLEKLLVYRAKEEGLNARREKIFFEIDELKQNNENLQSLSTEYS 787

Qy 668 LADGEHFIVDDKLSSTIVLEKILKLEGL-----DYSLRQTTQTFQFDMTGTAG 718  
Db 788 -----EELEKEKKILDELNEKIFTLRAEAVGNLLETKORYEKEMRDTG 829

Qy 719 KAL-----LSKSNLRQSGK--OKFQEAQAPFL-----GRV 746  
Db 830 KMIERIAREFTEDIKQMTSLEEMENYRKFIREFHERIEHLKXEMDSVFEMKLRSGKE 889

Qy 747 DLDKAIKAEKALVTKKATKNGQLLERSINKAVLAYNNSAIKANVRLEKELDLTLGLV 806  
Db 890 EKMRLEQVENRMDELKEEK--ERLRNHLHQIDLALQELTRLKIANL-----LEEFSGNE 941

Qy 807 EGKGPLAQATMVQGVYLLKTPPLPYVYIGLVNVDKSGKLIY-----ALDMSDTIGE 859  
Db 942 EDVEELDEE-----KLEETIYRQIK--DLENKIKYLGPDVLTDAIDSEYKLE 985

Qy 860 GOKDAYGNPILNVDEDNE---GVHALAVATLADYEGI--DIKTILNSKLSQTSIRQVP 913  
Db 986 EYEE-----ILKQKEDLEBAKRLKEIEIKTDREAESLLFDVYQVRNESFNFISL- 1036

Qy 914 TAAYHRAGIF---QAIQNAARAEQQLP-----KPGTHSEKSS 948  
Db 1037 -----LFFGSGRLNIVSEAKSILDAGFEISIRKPGRRDQKLS 1074

RESULT 60  
A45866  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans  
C/Species: Streptococcus mutans  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Mar-2004  
C/Accession: A45866  
F/Honda, O.; Kato, C.; Kuramitsu, H.K.  
J. Gen. Microbiol. 136, 2099-2105, 1990  
A/Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl  
A/Reference number: A45866; MUID:91100958; PMID:2148600  
A/Accession: A45866  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1431 <HON>  
A/Cross-references: GB:M29296  
C/Keywords: glycosyltransferase; hexosyltransferase  
F/181-201/Domain: cpl repeat homology <CP1>  
F/1127-1146/Domain: cpl repeat homology <CP2>  
F/1192-1211/Domain: cpl repeat homology <CP3>  
F/1257-1276/Domain: cpl repeat homology <CP4>  
F/1277-1297/Domain: cpl repeat homology <CP5>  
F/1321-1340/Domain: cpl repeat homology <CP8>  
F/1341-1361/Domain: cpl repeat homology <CP6>  
F/1385-1404/Domain: cpl repeat homology <CP7>

Query Match 3.5%; Score 176; DB 2; Length 1431;  
Best Local Similarity 17.9%; Pred. No. 2.5;  
Matches 191; Conservative 180; Mismatches 401; Indels 294; Gaps 43;

Qy 1 MKKHLKTVALTITVSVVTHNQEVFSLVKPEILKQTOASSISGADYAESGSKKLKINE 60  
Db 12 VKGHVTVAVASGLITLGT-----ITLGSVSASAEQTQTS----- 46

Qy 61 TSGPVDVTDVDFSDKRTTPEKIKONLAKGPREQELKAVTENSEKQI-TSGSOLBQSK 119  
Db 47 -----DKVTVQKSEDDKAAESSSQTDAPKTKAQ-----TEQTAQSQANVADTSTITK 96

Qy 120 ESLSLNKTVPTSNWEICDFITKGNLTVGLSKSGVEKLSQTHLVLPQSAADGTQLIQVA 179  
Db 97 ETPSQNTTQANS-----DDKTVNTKSEBAQTSEERTKQSEB---AQTTASSQALTOA 147

Qy 180 SFATPPDKTKTAIAE-----YTISRAGENCEISQLDVGKEITL--- 215

Db 148 KAELEKQRTAAQENKPNVDLAAIPNVKQIDGKYIYIGSDGQPKKNFALTNNKVLVYFDK 207  
Qy 216 NEGEVFN--YLLKKVTIPTGKHIGODAFVNDKNAIEVNLPESETISDYAFH-----268  
Db 208 NTGALTDTSTQYQFKQ-----GLTKLNDVTPHQNIVNFEN--TSLETIDNYVTADSWVRP 260  
Qy 269 -----LALKQIDL-----PNNLKAIGELAPFNDQITGKLSLPRQMLRAERA 310  
Db 261 KDILKNGKTWTASSDRLPLMSWPDQKQIAYLNYMNQOGLG-----TGENY 310  
Qy 311 FKSNIHKITFEKNSLKVIGESFQNDLSQLM-----LPDLEKIESEAFPTGNP 360  
Db 311 TADSSQESLNLAQTVQVKIETKISQTOQTOWLRDIINSFVKTPQNWNSQTESDTSAGE- 369  
Qy 361 GDDHYNNRVVLTGSKGNPSGLATENTYNNPKSLWQESP--EIDYTKWLEEDFYQKNS 418  
Db 370 -KDHLOGGALLYSNDK-----TAVANSYRLLNRTPTSTQKPKYFEDN-----413  
Qy 419 VTGFSNKGKLGKVRKNKLEIPKHNGVITTEIGDNAFRNVDPNQKTLARKYDLEEVKLPST 478  
Db 414 -----SSGGVDFLAN-----DIDNSNPVQAEQLN-----439  
Qy 479 IRKIGAFAPQSNLKSFEASD--DLSEIKEGAFNNRIETLELKDGLVTIGDAAAPHNH 535  
Db 440 -----WLHYLMNYGSIIVANDPEANPDGVRVDAVDNVDNLQI-----ASDYLKA 484  
Qy 536 IYAVILPESVQIEGRSAFRONGANNLI FMGSKVKTILGEMAFPLSNRLEHLDLSEQQLTEI 595  
Db 485 HYGVDKSE-----KNAINHLSIL-----EAMSDNDPOY---NKOTKGAQL 521  
Qy 596 PVQAFSDNALKEVLLPASLKTIREEAFKQKHLKQLEVASALSIAFNALDNDGDG---E 651  
Db 522 FI-----DNKRLSLIYALTRPLEKASKN-----EIRSGLEPVITNSLNRSABGKNS 572  
Qy 652 QPDNKKVVKTHNSY--ALADGEHFTVDP--DKLSSTIVDLKILKLEGLDYSTLRQTT 707  
Db 573 RMANVIFTRAHDSEVQTVIAKIIKAQINPKTDGLTFTLDELQAKFIYNEDMRQAKKYT 632  
Qy 708 QTOPRDMTITAGKALLSK-----SNLRQGEKQKFLQEAQFPLGRVLDLDAKAKKALVTK 763  
Db 633 QS--NIPTAYALMISNKDSITRLYGD--MYSDDGQYMATKSPYDAIDTLKARI--K 685  
Qy 764 ATKQOLLERSI--NKAVLAYNNSAIKANKVRLKELDLTLGLVEGKPLAQATMVQ 820  
Db 686 YAAGQDMKITVEGDKSHMDWYTGVLTS-----VRYGTGANEATDQGEAKTKQG 737  
Qy 821 VYLLKTPPLPEYIYGLNVPYFDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDEGYH 880  
Db 738 MAVITSNRP-----SLKLNQNDKVIV-----NMGAAHKQEVYRPLLTTKDGLTSY 783  
Qy 881 ALAVATLADY-----EGLDKITLNSKLSQTSIROVPTAAVHRAGIQAIONA 929  
Db 784 TSDAAAKSLYRKTNBKGELVFPASDIQGLNPQVSGYLAV--WVPVGASNDQDVRVAASNK 842  
Qy 930 AABAEQOLLPKPGTHSEKSSSSPS-----ANSKDRGLQSNPKTNR 968  
Db 843 ANATQVY-----ESSSALDSQLIYEGFSNFQDFVTKDSYTNK 881

RESULT 61  
B97116  
chromosome segregation SMC protein, ATPase [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: B97116  
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: B97116  
A:Status: preliminary

RESULT 62

A:Molecule type: DNA  
A:Residues: 1-1191 <KUR>  
A:Cross-references: UNIPROT:Q971A1; GB:AE001437; PIDN:BAK79717.1; PID:g15024720; GSPDR:G  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC1751

Query Match 3.5%; Score 175.5; DB 2; Length 1191;  
Best Local Similarity 19.2%; Pred. No. 2;  
Matches 170; Conservative 154; Mismatches 316; Indels 245; Gaps 42;

Qy 29 KEPI-LKQTQASSISGADYAESGSKLKNINETSQVDDTVTLDFSKRTTPEKIKONL 87  
Db 262 KEKISLELKVATESIDEPDAKYSSNKTXYESKSHQKILSEIIELEKTSNSDVAKNKL 321  
Qy 88 -----AKGPREQELKAVTENTSEKQITSG-SOLEQSKESL-----SLNKTIV 128  
Db 322 YKEIEDLNSIVNLKRYEIQLTLTEDKNYNKSKKSEKKKNIDGLIEBEWKS 381  
Qy 129 PSTSN--WEICDFITKGNLTGLSGVGEKLSQTDHLVLPQAAADG-----172  
Db 382 KQYKNDADIIISTISQNNNEVVILKEIE--SNESKLESIKRAGEGYSKSLKINEVTNT 439  
Qy 173 --TOLIQVASPAFTPKKTAIAEYTSRAGEN-GEISQLDVPDQKEIINBGEVFNYSLLK- 228  
Db 440 LSEELVKI-----NDK--ISGYENQIRENRSKISKLN---RIISDEEKLNNRELNSKS 486  
Qy 229 -----VTIPTGY-----KHIGQDAFVNDKNIAE-----VNLPESETI 261  
Db 487 NKLEANKMMLINLEKQYEGYNRSVKNLMQHVTK-GFVDVKPSESSFVLGEVKKVKEFE 545  
Qy 262 SDYAFALHALAQIDLPNL-----KAIGELAPD-NOITG-KLSLPRQLMRLAER 309  
Db 546 VEISIGAASIDIITDDNIAKLNLYKSNLGRATFLPLNIIKGRKLNISDATH--EK 603  
Qy 310 AFKSNHIKTIIFRGNLSK-----VICEASFQNDLSQLMLPDLCKIESEAFTPGNPDH 365  
Db 604 GFIGIASLIDYDSTFLPAVNVGLRTVI VDNMSALK-----IAKLSN-----Y 648  
Qy 366 NNRVVLWTKSGNSGLATENTYNNPKSLWQESPEIDYTKWLEEDFYQKNSVTFGPNK 425  
Db 649 SFKIVTLTGEVNVFGSITGSGTYSKAAASITGRKEI-----EELNLELNNVQALEQ 701  
Qy 426 GLQKVRKNKLEIPKHNGVITTEIGDNAFRNVDPNQKTLARKYDLEEVKLPSTIRKIGAF 485  
Db 702 SSNKIENKV-----VKEL-DMLCLDL---TDTIHGEKIELTKIKERLKSID-- 745  
Qy 486 AFQSNL--KSFEAS--DDLEBIKEGAFNNRIETLELKDGLVTIGDAAAPHNHIAVLPE 543  
Db 746 -IESEKLNKSYNTSVEIGFIKEK--INKHLEKLVKE-----780  
Qy 544 SVQIEGRSAFRONGANNLI FMGSKVKTILGEMAFPLSNRLEHLDLSEQQLTEIPVQAFS-- 601  
Db 781 --EENKALKLREANNNLI-----DELERKLDKNSKVLNLEINSMK 822  
Qy 602 -DNALKEVLLPASLKTIREEAFKQKHLKQLEVASALSIAFNALDNDGDGQFQNDKVVVK 660  
Db 823 VDKAKSDEMLMSSTREIERYKVEHNMENKISLK-----NEIDD-----FENQVKIL 870  
Qy 661 THNSYALADGEHFTVDPDKLSSTIVDLKILKLEGLDYSTLRQTTOTQPRDMTTAGKA 720  
Db 871 KNKLEYN-----NGKVQIKQVIELE-----NSFKDSEVERSK 904  
Qy 721 LLSKSNLRQGEKQKFLQEAQFPLGRVLDK-----AJAKKALVTKKATKNGQLLERSI 775  
Db 905 L--KGNIE--NKRNSLOQINLVLOKLETEKHRYEINLAKIE-----TESENLYERLN 952  
Qy 776 NKAVLAYNNSAIKANKVRL--EKELDLLTGLVEGKPLAQATM 817  
Db 953 SDFKLYTSEAAAEFKEIEEMISYKKCIDELKREINKMGVNVVASI 997



Db 470 ----NINQOIFKTEYLPFFE-----KGKEQAKLDYGNILNPYNTQAKVEALPK 517  
QY 358 GNPQGDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKN 417  
Db 518 GNKNQEIY-----QALDGNVAYEPG 537  
QY 418 SVTGSNKGLOKVRKNLEIPKQHNQVITITIGDNAPRNVDFONKTLKRYDLEEVKLP 477  
Db 538 AFKSVLNSWTGKIQHPEKADIQRFTHLEQVQKIGNSVLN-----QPO 580  
QY 478 TIRKIGAFQFQSNLKSPEASDDLEIEKEGAFNNRIETLELKDKLVITIGDAFHNHIY 537  
Db 581 TTKQVIVSLKNNF-----FKNGHVOVASYFQDLLTKDKL-TVLETLYDLAKKW 628  
QY 538 AL-----VLPEVQIEGRSAFRONGANNLI FMGSVKVTIGEMAFLSNRLHLDLSEOKOL 592  
Db 629 GLETNWAQPKAFQYTKDIFAB--ADKLKFLWKKK-----DPYNOI 669  
QY 593 TEIPVQAP-----SDNALKEVLLPASLKTIRREAFKQNHKQLEVASALSHIAPN 642  
Db 670 NEIHQLSFNILARNDVIKSDGYGVLLLPQSVKT---ELEGKNEAQ-----IFE 715  
QY 643 ALDDNDGDEQFNKVVVTHNSYALADGEHEFIVDPDKLSSTIVDLKILKILIEGLDYST 702  
Db 716 AL-----KKYSLIENSAP-----KTTILD-----KNLEGTDFKT 745  
QY 703 LRQTTQTQPRDMTTAGKALLSKNSLRQGEKQKQFLOBAQF--FLGRVLDLDKATAKAEKALV 760  
Db 746 -----FGDPL-----KAFFLKAAQFNFPAPWAKLDDNLQYSEAIK 781  
QY 761 TKKATNGQLLRSINKAVLAYNNSAIKKNYKRLKELD-LLTGLV-----EGKGP 811  
Db 782 KGETTKEG-----KREEDVKVKELDNKIKGILPQPPAAKPEAAKP 822  
QY 812 LAQATWQGVYLLKTPPLPEYIYGLNVYFDSKGLIYAL-DMSDTIG--EGQKDAYGN 867  
Db 823 VAAKPVAAKPEAAKPVAAKPE-----AAKPVAAKPEAAKPVAAKPEAAKPVAAK 871  
QY 868 PLIN-----VDEDNGYHALAVATLADY---EGLDIKT-----TLNSKL 903  
Db 872 PVATNTNTNTGFSLTKPKEDYFPMAFSYKLEYTDENKLSKTPEINFLVLELHQSEY 931  
QY 904 SOLTSIRQVPTAAYHRAGIFQIONAAAEOLLPKPGTHSGSSSESANSKDRGLQSN 963  
Db 932 EDOKIIEKDKTVNLQYQFQVVKVTSQYQKLSHPMTTEGTQNGQKKGEGFPNGKKAE 991  
QY 964 PKTNGRHSAILPRTGSK 981  
Db 992 GAPSQGKKAEGTFNQKK 1009

RESULT 64  
AG1717  
probable peptidoglycan bound protein (LPXTG motif) lin2283 [imported] - Listeria innocua  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AG1717  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AG1717  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1151 <GLA>  
A:Cross-references: UNIPROT:Q929J2; GB:AL592022; PIDN:CAC97511.1; PID:g16414795; GSPDB:C  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lin2283

Query Match 3.4%; Score 175; DB 2; Length 1151;  
Best Local Similarity 18.9%; Pred. No. 2;  
Matches 234; Conservative 171; Mismatches 408; Indels 424; Gaps 54;

QY 8 VALTITVSVVTHNQE-VFSLVKEPILKQTOASSISGA--DYAESGSKSLKINS-TSG 63  
Db 102 IYLVKIGNTIYKNAKLVINLPKNGEFTQSLSELKAGVTPSYNKTTRQLRVVLYKYOTLNSG 161  
QY 64 PVDDTVTDLFSDKRTTPEKIK-----DNLAGPREQELKATENTESEKQITSGSQ 114  
Db 162 VVDKVLKLSKNGYTPNGTKLEVTGBFSADNL-----EKVTEQAETTVNATATTA 213  
QY 115 L--EOSKESLSLNTKVPSTSNWEICDFITKGNLTIVGLSGSVEKLSQTDHLVLPQAAADG 172  
Db 214 LSNEFTKVENSIINN-PSQGDIGISF-----NLNIPKSTGS 250  
QY 173 TQLIQVAFAPTPDKKTAIAEY-----SRANGEISQLDVGKBI-----IN 216  
Db 251 -----LFIEGKIIIEYTLADKLDYLGAGDTPEPTK--IEGQKLWEIAAPTLY 299  
QY 217 EGEVFNYSLLKKVTIPTGYK-HIGQDAFVDNKNIAEVLNLPESLETISDYAFALHALKQID 275  
Db 300 BOEKATSLNKTFFQIRTFQTSIPNFATVENKAVATTNFTITLSDSIVDTSKASVSASD 359  
QY 276 LPDNLKATGELAFDQITGKLSPLQMLRAERAFKSNHIKTIIEFRGNSLKVIGEASQ 335  
Db 360 PATIPPTIGSV-----YAPAHRCPVDENWGIAVTGPNPDK 395  
QY 336 DNDLSQMLPDLGLEKIESEAFGNPGDD--HYNNRVVLWTKSGKNPSGLATENTYVNPDK 393  
Db 396 AYDSAKL-----GFSMLNSAMNDSPSYFIYV-----DVYVNIID 431  
QY 394 SL-----WQESPEIDYTKWL-----BEDFTYQKNSVTGFNKGLO 428  
Db 432 NLNLDYFRSGDYFYKPNVNYPGWPQPKKPKYNLLVKYDGDNDMTLTKEDVE-----LS 485  
QY 429 KVKRNKNIIEIPKQHNQVITIBI-----GDNAFRNVDFQNTLRYKDYLEEVKL 475  
Db 486 KMYSEKDLGIDDDKH--VSKVWLHFTYAPAGMYSADLSFFT-----TVKEGYGVEVRN 536  
QY 476 PSTIRKIGAFQSNLKSFEASDDLEE-----IKEGAFMNNRIETLE 518  
Db 537 TTKINMYGADS--QGYIHYFDDTNPWPENWOKYAGDRTAQIIPQPIGKNKFVQGVAPDD 594  
QY 519 LKDKLVITIGDAAFHNHIAIYVLVESVQIEGRSAFRONG--ANNLI FMGSVKVILGEMA 575  
Db 595 TDGNLINIGDNSVSVN-----LESNKASISRLTGPFEAMVLLPSGVMMKNTDQOG 644  
QY 576 FLSNRLEHLDLSEOKLQTEIPVOAFSDNALKEVLLPASLKTIR----- 618  
Db 645 FNVSVLNSNYQNSGRQLLKVKWD-----KKTLLPAEKLTAKNVTVSKDTPSNMVBEM 697  
QY 619 -----EEA-----FKQNHKL-- 628  
Db 698 FGFLQDQDFNVPEVTGTPTISDTKMEIDSNINDONGNSESRITSGNHVILNTSNHLKIS 757  
QY 629 -----OLEVASALSHIAPNALDDNDGEQFQPNKVVVTHNSYALADGEHIV 676  
Db 758 KKAAGNLDKGYSGLANATTNSIVSYQLLENDSEKIANMVLMDVLPSEKDLG-----IT 812  
QY 677 DPDKLSSTI-VDLKILKL-----IEGLDYSTLRQTTQTFRD-----MTTAGKALLSK 724  
Db 813 DNSRSGKFNLELTKAVAI PNWDKVB-VYTSKAKNPKRSGVLVDKHTIYPTGSPFLVDN 871  
QY 725 SN-----LRQGEKQKFLQEAQFFLGRVLDLDAI-----AKAEKALVTKATKNGQLLRS 774  
Db 872 REATEADWLTAQEVQDWSKIHSP---XWELKEGIEWIPGKSMKIQFDLTKPKQIDKEL 928  
QY 775 INK-----AVLAYNNSAIKKN-----VKRLSEKL---DLL 802  
Db 929 LNQTKTKEDRAAWNSFAVANVNSQVIEPAGVGVALDDSVAPVTVQVYDQNHKQIASPETL 988



QY 190 A--IAEYTSRAGE--NGEISQDVGKEIIINEGEVNSYLLK-----KVITPTGYKHI--- 238  
DB 576 LDHIGKVTSEHEVPKRVADADQAKSVKTKKSRKAKTPAKEDTLVDGSAQNVPEPI 635  
QY 239 -----GQDAFVDN-----KNIAEVNLPESLETISDYAFALHAKQIDLPDNLKAIGE 285  
DB 636 KVVDCGHDNVIRNLDLSIQORNEAENMEKSGKSS-----KRSKKDLSNIVEE 686  
QY 286 LAFFNOITGKLSLPRQLMRLAERAFKSNHIIKTIERGNSLKVIGEA---SFQDNDLSQL 342  
DB 687 AQVUS--LOQKEAENLEKSGKSKSKK-----DSLNIVEAQVLSVEVNVNQ- 738  
QY 343 MLPDGLKEKTESAFTGNPDGDHYNRVLWTKSGKNPSGLA-----TENTY-VNPD 392  
DB 739 -----EASPINPKD-----ALFPAKNTESNASPLKKIIEVTDINTEDINRS 784  
QY 393 KSLWESPEIDYTKLEEDFTYQKNSVTGFSNK-----GLQKVRKNKLEIPKOHNGVT 446  
DB 785 MQVQKENAGMG-----DNIGSSQKODIVGCANKQDQVGTGKSKKESLIDL---HPGGS 836  
QY 447 ITSIGNAFRNVDQFNKTLKYDLEEVKLPSTIRKIGAPAFOSNLSKFEASD-----D 500  
DB 837 ID-----GSMKMKETKGRVQPSSTGTSQLOSNMKNDRSGSKVD 874  
QY 501 LEEIKEGAFMNNRIETLEKDLKLVITGDAAFHNNHIAIYVLPESVQEIGRSAFRONGANN 560  
DB 875 LSDAPMKGTVNNKEAVKSSKSVTVNKSMMVNN-----KEKAVKISNSVTANKSTTN 929  
QY 561 LIFMGSKVKTGEMAFPLSNRLEHLDLSEQKQTEIPVQAFSDN 603  
DB 930 -----PFKDAEEDSKTTSSTDSTKAPSDSSDN 957

RESULT 67  
T13030  
microtubule binding protein D-ClpP-190 - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: T13030  
R:Lantz, V.A.; Miller, K.G.  
J. Cell Biol. 140, 897-910, 1998  
A:Title: A class VI unconventional myosin is associated with a homologue of a microtubul  
A:Reference number: Z17588; MUID:98139549; PMID:9472041  
A:Accession: T13030  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1690 <LAN>  
A:Cross-references: EMBL:AF041382; NID:g2773362; PID:g2773363; PIDN:AAB96783.1  
A:Experimental source: strain Oregon R  
C:Genetics:  
A:Cross-references: FlyBase:FBgn0020503  
C:Keywords: cytoskeleton

Query Match 3.4%; Score 174.5; DB 2; Length 1690;  
Best Local Similarity 19.7%; Pred. No. 3.7;  
Matches 244; Conservative 189; Mismatches 448; Indels 359; Gaps 54;

QY 3 KHLKTVALLTVTVSVVTHN-----QEVSLVKPEILKQTOASSISGADYAESGK 53  
DB 460 KITLVSATPSILPDLPSDDGALQBEIAQLQEKMTIQKEVESRIAEQLREQLR 519  
QY 54 SKLK-INETSFPVD-----DVTVDLPSDKRTTPEKIKDNKAGPREQELKAVTENTESE 106  
DB 520 ENVKYLNEQIATQSELVSKDALKFSLSECGIENLRRELALLKEENEKQAEQAEFT 579  
QY 107 KQITSGS--QLQSQSKESISLNTKTPSTSNWEI-----CDFITKNTLVLSKSGVEKLS- 158  
DB 580 RKLAEKSEVLRSLSELQNLKATSDLSLSESRVKNKSDCEIL---QTEVMRDEQIRELNQ 636  
QY 159 QTDHLV--LPSQAADGTQIQVAFPT--PDKTIAEYTSRAGENISQLDVVDGKEII 215  
DB 637 QLDVTTQLNVQKADSSALDMLRLQKEGTESKSTLLEKTEK-----ELVQIKEQAANTL 691

QY 216 NEGEVNSYL--LKKVTTIPTGYKHIGODAFVNDKN---IAEVNL--PESLETISDYAFALH 269  
DB 692 QDKQLEKQISDLKOLA-----EQKLVREKTENAINQIQLEKESIE-----OOL 736  
QY 270 ALKQIDLPDNLKAIGELAFFDQNTGKLSLPRQLMRLAERAFKSNHIIKTIERF--RGNSLK 327  
DB 737 ALKQNELEDPOKQKQSESEVHLQEIKAQ-----NTQKDELVESGESLK 779  
QY 328 VIGEASFODNDLSQLMLPDGLEKTESAFTGNPDGDHYNRVV-----LWTKSGKN 378  
DB 780 KL--QQOLEEKTGLGHEKLOAALDELKCEKT-----IIKEKEQELOOQLOKSAES 827  
QY 379 PSGLATENTYNNP-----DKSLWESPEIDYTK--WLEBDFYQKNSVTGFSNK 425  
DB 828 ESALKVVOVQLEQLOQQAASGEGSKTVAKLHBEISQLKSAEQTSQELKSTENLEAK 887  
QY 426 GLQKVRKNKLEIPKOHNGVTITTEIGNAFRNVPQNKTLRKYDLEEVK--LPSTIRKIG 483  
DB 888 SKQLEAANGSLSEEAQKSGQLEQI-----TKLSEVEETOAAALSSVHTDV- 933  
QY 484 AFAPQSNWLSFEASDDLEEKEGAFMNNRIETLEKDLKLVITGDAAFHNNHIAIYVLP 543  
DB 934 -----ESKTKQLEAANAALAEKVNREYAESRAESDLODKVKEITDT-----LHAELQAE 982  
QY 544 ---SVQEIGRSAFRONGANNLIFMGSKVKTGEMAFPLSNRLEHLDLSEQKQTEIPVQ- 598  
DB 993 RSSSALHTKLSKP-----SDEIATGHKELTSKADAWSQEMLQ-----KEKELQELRQOL 1032  
QY 599 -----AFSDNALKEVLLPASLKTIREBAFK-----KNHLKQLEVA 633  
DB 1033 QDSQDSQTKLKAEGERKEKSPESIKNLQEEVTRAKTENLELSTGTQTTIKDQERLEIT 1092  
QY 634 SA--LSHTAFNALDNDGDGEQFDNKVV-----KTHH 663  
DB 1093 NAELOH--KEKQASDAQIADLKTVEAIQVANANISATNAELSTVLEVLQAEKSETNH 1150  
QY 664 -----NSVALADGEHFIVDPDKLSTIVDLEKILKILEGLD-----YSTL 703  
DB 1151 IFELFEMEDANNSERLIEKVTGI--KEELKETHIQLDERQKRFEELEBKIKQAQOSEOKL 1208  
QY 704 RQTTQTOFRDMMTAGAL-----LKSNNLRQ 729  
DB 1209 QESQTSKEKLETEIQOSIQELQDSVKQKEELVQNLEEKVRESSSIIIEAQTINKNESVQL 1268  
QY 730 GEKQKFOEAQFFLGRVDLDKAI--AKAEKALVTYKATNGOLLERSINKAVLAYNNSAI 787  
DB 1269 ENKTSCLKETO-----DQLLESQKKEKQLOBEAAKLSGEL--QQVQEAANGDIKDSL 1318  
QY 788 K-KANVKELEKELDLTLGLVEGKGLQAAT-----MWQGVYL----- 823  
DB 1319 KVESLVKVELEKQAATSQDQAQ-----QATNKEIQELLVKSQENEGNLQGSSELAUTEKLO 1374  
QY 824 -----LKTPLPLPEYIIGLVNVFVKSGKLIYALDMSDTIGECQKDAYGNPILNVDED 875  
DB 1375 QLEQANGELKEALCQKB-----NGLKELQK-----LDESNTVLESQKSHNEIQDKLSQA 1425  
QY 876 NEGTHALAVAT--LAD-----YEGLDIKTILNSKLQSILTS-----IROVP 913  
DB 1426 QOKERTLOEETSCLAQESQLKQANEELQKSLQOKLLEKGNBFDQLABYQKVIDEMD 1485  
QY 914 TAAVHRAGIFOAIQNAAEAEQLPKPQTHSEKSSSESAN-----SKORGLOSN 963  
DB 1486 DAASVALLQOLQNRVALEETAL--RQANDQKTAYLETKELRQLESLELEKREVLJSL 1544  
QY 964 PKTVRGHSAILPRTGSGSFVYIGILYTSVALLSLITAI 1003  
DB 1545 KAQWNGASS-----RSG-KGDEVESLIDITSLAKINFINSI 1579

RESULT 68  
B71603  
RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)









Db 585 LKEDYVSPLESLYLVHAKATNHHLLSDSAYEDLVKCKENPDMEFLKEKSAKLGHTVSN 644  
QY 386 NTVNPKSLWQSPIDYTKWLEEDFTYQKSVTGFSGKQKVKRKNLEIPKOHNGV 445  
Db 645 EAYSELEKKL--EOPSLVY--VEHAKATNHHLLSDSAYEDLVKCKENPDMEFLKEK-- 698  
QY 446 TITEIGDAPNRVDFONKTLRKY-DLE-EVKLPSTIRKIGAFAPQSNLKSFEASDDLRE 503  
Db 699 --AKLGHTVSN-EAYSELQRYKSELEKEVEQPSLAYLVHAKATDHHLLSDSAYEDLVK 755  
QY 504 IKEGAPMNRITELKDKLVITGDAAFHINHIYAVLPESVQETORSAPRONGANNLIF 563  
Db 756 CKE---NPDVEF--LKESAKLG--HTVVSSEYSELQRY----- 789  
QY 564 MGSVKVTLGEMAFSLNRLEHLLDSEKQLTEIPVQAFSDNALKEVLL---PASLKTIREE 620  
Db 790 --SELEKEVEQPSLAYLVHAKATDHHLL-----SDSAYEELVKCKENPDMEFLKEK 839  
QY 621 AFKQNH-----LKQ-----LEVASALSH--IAPNALDD-----NDGDEQF 653  
Db 840 SAKLGHTVVSNEAYSELEKKLEQPSLAYLVHAKATDHHLLSDSAYEDLVKCKENSDFEF 899  
QY 654 DNKVVVKTNN-----SYALADGEHFIVD-----P 678  
Db 900 LKEKSAKLGHTVVSNEAYSELEKKLEQPSLAYLVHAKATDHHLLSDSAYEDLVKCKENP 959  
QY 679 D-----KLSSTIVDLKILKLEGLDYSLRTTOTQPRDMTTAGKALLSKS---- 725  
Db 960 DMEFLKEKSAKLGHTVVSNEAYSELEKKLEQPSLEYLVE--HAKATNHHLLSDSAYED 1015  
QY 726 --NLROGEKQKFLQEAQFIFGR-VLDLKAIAKAELAVT-----KKATKNGQLLER 773  
Db 1016 LVKCKENPDMEFLKEKSAKLGHTVVSNEAYSELEKKLEQPSLEYLVHAKATNHHLLSDS 1075  
QY 774 SINKAVLAYNN---SAIKKANKV-----RLEKELD--LITGLVE-----GKQP 811  
Db 1076 AYSELVKCKENPDVEFLKEKSAKLGHTVVSNEAYSELEKKLEQPSLEYLVHAKATNHHLL 1135  
QY 812 LAQATWVQGVYLLKTLPLPEYIGLVYF--DKSGKLIYALDMSDTTIGEGOKDAYGNPI 869  
Db 1136 LSDSAYEELVKCKENP-----DVEFLKEKSAKLGHTVVSNEAYSELEKKLEQPSL 1185  
QY 870 LNVDEDNEG--YHALAVATLADY-----EGLDIK-----TILNLSKLSQTSIRQ 911  
Db 1186 AVIVEHAKATDHHLLSDSAYEDLVKCKENPDVEFLKEKSAKLGHTVVSNEAYSELEKKLE 1245  
QY 912 VPTAYHRAGIFQATQNAAEQLL 937  
Db 1246 QPSLAY-----LVEHAKATDHHLL 1264

RESULT 73  
T14156  
Kinesin-related protein - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T14156  
R:Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.  
Cell 91, 357-366, 1997  
A:Title: KENP-E is a plus end-directed kinetochore motor required for metaphase chromosome segregation  
A:Reference number: Z17893; MUID:98028574; PMID:9363944  
A:Accession: T14156  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2954 <WO>  
A:Cross-references: UNIPROT:O42263; EMBL:AF027728; NID:G2586071; PIDN:AA68  
C:Genetics:  
A:Gene: KENP-E  
C:Superfamily: centromere protein E; kinesin motor domain homology

Query Match 3.4%; Score 173.5; DB 2; Length 2954;  
Best Local Similarity 20.1%; Pred. No. 9.5;  
Matches 209; Conservative 186; Mismatches 431; Indels 213; Gaps 47;

QY 23 EVPSLVKBPILKQTOASSISGADYAESGSKKLKINETSGPVDVTVTLDFSDKRTTPBK 82  
Db 1200 EVKLLEWELLKHGHTDSQLS-----IEKQLENLEVTETKLTQLQEBMKNTITERNELQTN 1255  
QY 83 IKDNLAKPREQELKAVTENTSESEKQITSGSQLEQSKESLSLNKTVSPSTNNWEICDFITK 142  
Db 1256 PED--LKAHDSLKQDLSENI--EQSIETQDELRAAQBELREQQLVDSFRQQLD--- 1307  
QY 143 GNTLVGLSK-----SGVEKLSQTHLVLPQAAQDGTQ-LIQVAFAPTPDKKTAIAYT 195  
Db 1308 --CSVGISPNHDAVANQEKVSLGEVNSLOSEMRLGRDELQTSCKALVSELELRAHVK 1365  
QY 196 SRAGNNGIS--QLDVGKEIINEGEVFNLSYLLKVKVTIPTGYKHIGQDAFVDNKNIAEVN- 253  
Db 1366 SVGENLEITKGLNGLEKEILKSE--ESEVLKSM-----ENLKEDNN 1407  
QY 254 -LPSELETISD-----YAFALHALKQID-LPDNLKA-----IGELAFFD----- 290  
Db 1408 KLKEQAEVSSKENQFSLBEVFGSQKLVDEIEVLKAQLKAAERLEIKDRDYFELVOTA 1467  
QY 291 --NQITGKLSLPROLMRLAERAFKSNHIK-TIEPRGN--SLKVICEASFODNDL----- 339  
Db 1468 NTNIVEGKLETPQ-----ADHEEDSIDRSEEMEIKVLGEKLERNOYLLERLQE 1517  
QY 340 SOLMLPDGLKEKTESAFTGNPGDDHNNRVVVLWTKSGKNPSPGLATENTVV--NPDKSLWQ 397  
Db 1518 EKLELSNKLEILOKEMETS-----VLLKDDLOQKLSLSENILKENIDITLKH 1567  
QY 398 ESPEIDYTKWLEEDFTYQKSVTGFSGK--LQVKRNKNLEIPKOHNGVTITEIGDAP 455  
Db 1568 HSDTOAQLOKTOQLAKLAKLAIASDNCPIQEBKETSADCVHPLEEKILLITE---ELH 1624  
QY 456 RNVDQNTKL-RKYDLE--EVKLPSTIRKIGAFAPQS--NNLKSF--ASDDLEBEIKEGAPM 510  
Db 1625 OKTNEQKLEKHEKNEQAQVELKCEVHLMKSMIESKSSLESLOHEKHDTBQQLLAKQ 1684  
QY 511 NNRIETLEKDKLVITGDAAFHINHIYAVLPESVQETIGRSAFRONGANNLIFMGSVKVT 570  
Db 1685 QMVVTOQKLEKLEQTHLTAEDVH-----LKENI-EUGLN-FKNE-----AQQT 1728  
QY 571 LGEMAFSLNRLEHLLDSEKQLTEIP--VQAFSDNALKEVLLPASLKTIREEAFKKNHLK 628  
Db 1729 TKEQCLL--NENKLEQSQHRLQCEIEELMKSLKQ---KE---SALETKESEQKVINLN 1780  
QY 629 Q-----LEVASALSHIAFNALDDNGDSQDFDNKVVVKTNN-----SYALADGEHFIV 676  
Db 1781 QEMEMVLEMEBELKNSQRTVIAERDQDLDRESVEMSIETQDLDLRKAEALQOQKDKV- 1839  
QY 677 DPDKLSSTIVDLKILKLEGLDYSLRTTOTQPRDMTTAGKALLSKSNLRQGEKQKF- 735  
Db 1840 --QELTSQISVLQEKISLGE-----NQMLYNAVTKETLSEKDDLNQSKQHIFS 1886  
QY 736 -----LOBAQFFLGRVDLDKAIKAEKALVTKKATK--NGQLLERSINKAVLAYNNS- 785  
Db 1887 EITLSLSLKEKEFALEQAEKDAKARTIDITEKISNIEBQLLQOATNLKETLYERES 1946  
QY 786 -----AIKKANVVRLEKELDLTLGLVEGKPLAQATWVGQVYLLKTPPLPYITGL 837  
Db 1947 LIQCKEQLALNTEHLRETLKSKDLALGRMEQERDEA---ANKVIALTEKMSLSLEEQINE 2002  
QY 838 NVYFDKSG-----KLTY-----ALDMSDTTIGEGOKDAYGNPILNVDE-----DNEGYHA 881  
Db 2003 NVTTLKEGEGEKETYLQRPQKSSSQMEELRESLTKTDLQLEBAEKEISEATNEIKNL 2062  
QY 882 LAVATLADYEGIDIKTILNLSKLSQTSIRQVPTAYHRAGIFQATQNAAEQQLPKPG 941  
Db 2063 TAKISSLBEEILQNASILNEAVSERENLR-----HSKQQLVSELEQL--SUT 2107  
QY 942 THSEKSSSESANSKDRGL 960  
Db 2108 LKSRDHAPAQSKRENDEAV 2126



A;Accession: S54985  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-1332 <GAW>  
A;Cross-references: EMBL:L22537; NID:g349189; PIDN:AAC37424.1; PID:g349190  
R;van der Aart, Q.J.M. Data Library, August 1995  
submitted to the EMBL Data Library, August 1995  
A;Reference number: S59702  
A;Accession: S59716  
A;Molecule type: DNA  
A;Residues: 1-1332 <VAV>  
A;Cross-references: EMBL:X76294; NID:g974203; PIDN:CAA53940.1; PID:e264674; PID:g558360  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: SGD:SPT7  
A;Cross-references: SGD:S0000285; MIPS:YBR081C  
A;Map position: 2R  
C;Superfamily: bromodomain homology  
C;Keywords: nucleus; transcription regulation  
F;466-521/Domain: bromodomain homology <BRO>

Query Match 3.4%; Score 173; DB 2; Length 1332;  
Best Local Similarity 19.7%; Pred. No. 3.1; Indels 248; Gaps 38;  
Matches 165; Conservative 130; Mismatches 296; Indels 248; Gaps 38;

QY 19 THNQEVSLVKEPILKQTOASSISGADYAESGSKLKN-----ETSGPVDVT 69  
DB 615 TNDKLTSLVPEGEKEDKTASST--VTVHENVNKNKEIKNGKNERQDMVESKTEDSS 672

QY 70 TDLFSKRTTPEKIKONLAK-----GPREQLKAVTENTSEKQITSGSOLEBSKE----- 120  
DB 673 KQDAAKKOTEDGLODKTAENKEAGENNEEDDDDEDDMDVDSQSYLLEKDDDRDDL 732

QY 121 SLSLNKTVPSTSNWEIC---DPIITKG--NTLVGLSKSGVEKLSQTDHLVLPQAAAGTQ 174  
DB 733 EISVWKTIVTAKVRAEICLKRTEYFKNGKLNDSSEAFKNPQRMKRPDQLFLEYK----- 786

QY 175 LIQVSAFPTPKKTAIAEYTSRAGENG-----ETSQLDVGDKETLINEGEVFP 221  
DB 787 -----EQKALESVROKIEQNSIMKNGFGTVLKQEDDDQLQFHNHSLNGNEAF 834

QY 222 -----NSVLLKKVTIPYCYKHIGQDAFVNDKNI---AEVNLPSLETISDYAFAH 268  
DB 835 EKQPNDEILDTRFLOEYDISNAIPDIVYEG-VNTKTLQKMEDASVDRMLQNGINKQSRF 893

QY 269 LALKQIDLP-----DNLKAIGELAFDQITGKLSLQRLM-----PLAERAPKSNHI 316  
DB 894 LANKDGLTQKMNQNTLITQQI-----RHICHKISLIRMLQSPLSAQNRSNPNFALNNHI 949

QY 317 KTIPEFGNSLKVIGEASFQNDLSQLMPLDGLKIESEAFTPGNPDHNNRVVLTGSG 376  
DB 950 YNYTIIDSLDI-----DPVSQJ-----PTHYKKNRELKWFVH 984

QY 377 KNPGLATENTYVNPDKSLWQESPEI--DYTKWLEEDFTYQKNSVTGFSNKGLOKVRNK 434  
DB 985 KNISKVAMANGPETAHPSAINMLTEIAGDYLSNLK--TLKLHHTNSLNRG-----T 1035

QY 435 NLHFKPHNGVITTEIG---DNAPRV--DFQNKTLRYDLEEVKLPSTIRKIGAPAQ 488  
DB 1036 NVBMLQ-----TTLLENGINRPDLFSYVESEFGKTKKLQDIKQ-KLESFLRALLRPTLQ 1090

QY 489 SNNLKSFEASDDLEETKEGAFMNNRIETLELKLVTIGDAAPHINHIYAILVPESVQEI 548  
DB 1091 ELBERNFE--DESQSPFTGDPASE-----LTGEDFP----- 1119

QY 549 GRGAFRONGANN-----LIFMGSKVKTL-GEAFLSNRL--EHLDLSEQOLTEI 595  
DB 1120 --GPRELGLEKEFGVLSVPLQLLTQFTQVDGETKVQAKKIQPEESDSIVYKIKTG 1176

QY 596 PQVAFSDNALKEVILLPASLKTIREAPKKNHLKQLEVASALSIAFNALDDNDGBOFN 655  
DB 1177 MLDA--GSFWNTLLP-----LLQKDYERSKAYIAKQSKSAN-----DK 1213

QY 656 KVVVKTHNSYALADGEHFI-----VDP-DKLSSTI-----VDLEKIL 692  
DB 1214 TSMTSTEDNSFALLEEDQFVSKTKATKARLPPTGKISTTYKKKPIASAFILPEEDLENDV 1273

QY 693 KLIEGLDYSLRQTTQTPRDMTTAGKALLSKNSLRQGEKQKFLQEAQFPLGRVLDKA 751  
DB 1274 K-----ADPTTTVNAKVGAEN--DGDSSSLFLRTPQ-PLDPLDMDDA 1311

RESULT 76  
D97033  
uncharacterized protein, probably surface-located [imported] - Clostridium acetobutylic  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: D97033  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: D97033  
A;Status: Preliminary  
A;Molecule type: DNA  
A;Residues: 1-1819 <KUR>  
A;Cross-references: UNIPROT:Q97K40; GB:AE001437; PIDN:AAK79055.1; PID:g15023996; GSPDB:  
C;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC1081

Query Match 3.4%; Score 173; DB 2; Length 1819;  
Best Local Similarity 18.3%; Pred. No. 4.9;  
Matches 206; Conservative 186; Mismatches 388; Indels 348; Gaps 52;

QY 38 ASSSISGADYAESGSKLKNETSGPVDVTYDLPSDKRTTPEKIKONLAKGPREQ-EL 96  
DB 36 SSKKIVTSEKASESGNNAVNNEIK-----TPDNSYEQAKNLQVQADV 79

QY 97 KAVTENTSEKQITSGSOLESKESLSLNKTVPTSTSNVEICDFITKNTLVGLSGSVBK 156  
DB 80 KQVQGVTVNTQ-TSNETDEKLINASAGENLSDYNSLGF-----TNVKPDNIKV 131

QY 157 LSQTDHLVLPQAAQDTQLIQVSAFPTD-----KKTAAEYTSRAGENGES 205  
DB 132 MBQIKNARTSGADLTQDEIGTTTANTSDSIGIFDRMTKGTAVVSDYQFLL-----IT 186

QY 206 QLDVVGKELINEGEVFNVSLLKKVTIPYCYKHIGQDAFVNDKNIYAEVNLPSLETISDYA 265  
DB 187 QVNSNMLQDINSWLTGKRYALISKILST-----VDKITTAVSNINNANDHESDFT 236

QY 266 FAHLALKQIDLPDLNLKAIGELAFDQITGKLSLQRLMLAERAPKSNHIKTIPEFGNS 325  
DB 237 ALQIYVQEDL-----VFFLNENIS-----AAKN-----KQSD 265

QY 326 LKVICEASFQNDLSQLMPLDGLKIE-----ESEAFTPGNPDHNNRVVLTGSKGN 378  
DB 266 LSSSEINTVWQDSLKLQ--DALERINLQOGLDDYHFIGAANVDENLLEDVNGWANGK- 322

QY 379 PSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVK----- 431  
DB 323 --GWFERNAIBEINSIVEPLSRINSQDSTAAD--YDAIQVDGVDGDDNNIDNIKSVIAEKK 378

QY 432 --RNKNLEIPK-----OHNGVTITEIGDAFNPRNDFONK 463  
DB 379 QLKGNLMAIEIKSAVEYKTVLDYFDLKKGTAKVSDYNAIGLTVTED---NVTDMNE 435

QY 464 TLRYKDLBEV-KLPSTIRKI-----GAFQSNLKS 495  
DB 436 LLKNRDITLNLKQDNINSIIKSLKNINAGTDPDYYTNLNTSSVTQDNISFIRDIKTA 495

QY 496 EASD--DLEETKEGAFMNNRIETLELKLVTIGDAAPHINHIYAILVPESVQETGRSAF 553  
DB 496 KOAGSDLNKKGKIQDSVNSNKLRLSAMDRI-NAGSAAALDDYNNLGI---EGVT----- 544





hypothetical protein C56E6.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T15864  
R:Fulton, L.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid C56E6.  
A:Reference number: S69019  
A:Accession: T15864  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1066 <FUT>  
A:Cross-references: UNIPROT:Q18902; EMBL:U39996; NID:g1055114; PID:g1055120; PIDN:AAA810  
C:Genetics:  
A:Gene: C56E6.6  
A:Introns: 62/3; 110/3; 376/3; 432/3; 482/3; 574/2; 649/2; 679/2; 769/2; 833/3; 892/3; 9

Query Match 3.4%; Score 172; DB 2; Length 1066;  
Best Local Similarity 20.1%; Pred. No. 2.5;  
Matches 162; Conservative 129; Mismatches 300; Indels 214; Gaps 39;

QY 5 LKTVLTLTVSVVTHNQEVSLVKEPILKQTOASSISGADYAESGSKLKNETS 64  
DB 212 LKNFAKTLKTL-----KLAENNIHATPEALRDLRLNLTHNLNGN---KLNRI 258

QY 65 V-----DDTVTDLFSDKRTTPEKIKDLAKGPRQELKAVTENTSEKQITSGSQESKE 120  
DB 259 VLKGCDDTLVELFIANNYL-EHPHGVLSQMKQLEHLDISKV-----KIMS---LKKPTS 309

QY 121 SLGLNTKVPST-----SNWEICDFI-----TKGNTLVGLSKSGVEKLSQ 159  
DB 310 LLSITKEETSTVRLNLNAGNRINNMDSYLIFFEHMPLLTIVDVSNFIRFISPRVPEKLN 369

QY 160 TDHLVLPSQ-----AADGTQLIQVASFAPDPKTAIAEYTSRAGENG 202  
DB 370 LESLFLQNNQLAHFSLFRDLKRLHMLDNNQIKDINFSLADLPK---LOHLSLAG--- 423

QY 203 ETSQLDVDGKEIINEGEVFNLSYLLKKVTPTGKVK-IGQDAFVDNKNIAEVNLP-ESLET 260  
DB 424 --NQLDIITENMGSS---SSSELKSLNLAHNKIHSISRSFSDLDNLQQLRLSHNNIIT 478

QY 261 ISDYAFAPHLA-LKQIDLPNLKAIGELAPFDNQITGKLSLQJMLAERAFKSNHIKT 319  
DB 479 ITSMTFSNLRNRYLDLSH-----RIIKLPSALYQLPALDVL 517

QY 320 EFRGNSLVIGESFOD-NDLSQMLPDGLEKIESEAFNPGDHDHNNRVVLWTKSGKN 378  
DB 518 HLDHNNLNEIDRDAFRSFDLSQSLKLSHNAFRFSCFGLSGISQVHQ----- 564

QY 379 PSGLAFTVNPDKSLWQESPEID--YTKMLEEDFTYQKNSVTGFSNKGLOKVKRKN 435  
DB 565 ---LDLSSNQIN-----EIDIFCARGIRK-LSLASNSVEKINRKLQDATELTS 610

QY 436 LEIPKOHNGVTITEIGNAF-----RNVDQFNKTLRK-YDLEEVKLPSTIRKIGAFAP- 487  
DB 611 IDI--SHNG--IIDVSDAFCECRKLSHKLSHNYIRNLWKTRVCIP-WISHLITFCFF 665

QY 488 -----QSNNLKSF-----ASDLEIEKEGAFNN-NRIETLELKOXL 523  
DB 666 TKHLELRTISFSYIIVDSSQLTSFGNLSILSPANNKVDSDIEGAFENLLSLKILDLSNPN 725

QY 524 VTIGDAAFHNIHIAIVLPESVQEGRSAPRQNGANNLIIFMGSKV--KTLGEMAFLSNRL 581  
DB 726 VTWS-----PTAFRDLSHSISINMANTGLFSMPKFSHSIQSLNISCNKI 772

QY 582 EHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREBAFKKHLKQLEV-ASALSHIA 640  
DB 773 YELSEKDLAPLTKVVALDISHNNLKQISSNAPEPLI-----HLKQLNVSNPITHLT 824

QY 641 FNALDDNDGDEQPDNKVVVTKTHNSYALADGEHFIVDDPKLSTIVDLKILKLEGLDY 700  
DB 825 -----NEHIQQL-----YKLNIPD-----MARPYQISSILSNLPLHTIYVDIKE 865

QY 701 STL-RQ--TTOTQ-FRDMTTAGKAL 721  
DB 866 SALDRQFYTADTFLRLHLVWAGN 890

RESULT 79  
A54817  
ATPase ScfII, chromosomal scaffold - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 12-Apr-1995 #sequence\_revision 12-Apr-1995 #text\_change 09-Jul-2004  
C:Accession: A54817  
R:Saitoh, N.; Goldberg, I.G.; Wood, E.R.; Earnshaw, W.C.  
J. Cell Biol. 127, 303-318, 1994  
A:Title: ScfII: an abundant chromosome scaffold protein is a member of a family of putative  
A:Reference number: A54817; MUID:95014725; PMID:7929577  
A:Accession: A54817  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1189 <SAI>  
A:Cross-references: UNIPROT:Q90988; GB:X80792; NID:g572691; PIDN:CAA56767.1; PID:g572692  
C:Superfamily: chromosome segregation protein SMC1  
C:Keywords: nucleotide binding; P-loop  
F:32-39/Region: nucleotide-binding motif A (P-loop)

Query Match 3.4%; Score 172; DB 2; Length 1189;  
Best Local Similarity 19.7%; Pred. No. 2.9;  
Matches 179; Conservative 130; Mismatches 325; Indels 274; Gaps 39;

QY 36 TOASSISGADYAESGSK-LKINETGPDVDTVDLFSDKRTTPEKIKDLAKGPRQ 94  
DB 414 TEAKQAKMLKYAQQLKTKQAEVKKMGDSYKED-QEAFAEAIKTKELQDEMCK----- 467

QY 95 ELKAVTENTSEKQITSGSQESLSLNKTVSTSNWEICDFITKGNTLVGLSKSGV 154  
DB 468 -LK--YEEAEQEAHLAKKKQLSSSEISLR-----ELCEIAKHPYLFYKPNP 513

QY 155 EK---LSQTDHLVLPSSQAADGTQLIQVASFAPDPKTAIAEYTSRAGE---NGEISQLD 208  
DB 514 EKNWNPNCVKGLV-----TLTVKDIS-----TSKALEAVAGGKLYNIV 553

QY 209 VD---GKEIINEGEVFNLSYLLKKVTPTGKVKHIGQDAFVDNKNIAEVNLPESLET 261  
DB 554 VDTGATGKILEGKQKHYRTIIPLSKISA---NSIGHEIISLAKNLI----- 598

QY 262 SDYAFAPHLAKQIDLPNLKAIGELAF-----FDNQITGKLSLQJMLAERAFKSN 313  
DB 599 -GHRVHIAISLDYNSSELQKAMEYVFGTTLVSSMDN--AKKVTDFDKRIM- 647

QY 314 NHIKTIEFRGNSLVIGESFODNDLSQMLPDGLEKIESEAFNPGDHDH---NNRVV 370  
DB 648 ---KTVTLQGDIPDQGTLS---GGASSHVP-----ILSKLTVRDADELKIKTSQLE 696

QY 371 LWTSGKNPSGLATENTYVNPDKSLWQESPEIDYTKMLEEDFTYQKNSVTGFSNKGLOKV 430  
DB 697 ATEKELANLKNMAEYQHLKQWQESKEAEALLQTKIQSSAVHKQEDL----- 745

QY 431 KKNKLEIPKQHNGVTITEIGNAFNVDVFNQNLKRVKLDLEVKLPSTIRKIG---AFAP 487  
DB 746 -----LALKK-----TIAACEETLKTESQKAEYKALENKNKAEERKEIKNAQ 795

QY 488 OSNNLKSFEASDLEIEKEGAFNNRIETLEKDLKLVITIGDAAFHNIHIA---IVLPE 543  
DB 796 QKLSAKKKADDSSRMKE---KQEVBALVLELQLEKQESYKQSSAAQQAASLKE 852

QY 544 SVQIEGRSAFR-----QNGANNLI-----FMGSKVKTL-----GEMAFLSNR 580  
DB 853 QVSALAEAVKTRSLKNAENLSSEKGLMEERTKDIKAKSAKIEKYREQNNELQSLNA 912

QY 581 LEHLDSLQKQLTEIPVQAFSDNALKEVLLPASLKTIREBA-----FKKHLKQL----- 630  
DB 913 LEH-DINKYQGT-ADASTLDLKLKEYKWIASEKLFQADTTTYDFEANNPKETGQKLQ 970

QY 631 -----EVASALSHIAFNALDDNDGDEQFDNKVVVTKTHNSYALADGEHFIVDDPKLSS 683

```

Db      971 KLTTKKLEKLSNMRANLL--SEAEERYNDLMKKR-----MVENDKI-- 1013
Qy      684 TIVDLKILKLEGLD-----YSTLRQTOTQPRDMTTAGKALLSKSNLRQGBKQ 733
Db      1014 -----KILATIEELDRKNKALHIAWEKVNKDFGIFSMLLPAGAKAMLVPS-----KKQ 1062
Qy      734 KFLQEAQPFPLGRVVDLKAIAKAEKALVTTKATKQGLLERSINKAVLAYNNSAIKKANVK 793
Db      1063 NLLDGLFRVGLDITWK----- 1079
Qy      794 RLEKELDLTLGLVEGKGPLAQATMVQGVYLLKTPLPPEYIYGLNVYFDKSGKLIYALDM 853
Db      1080 -----ENITELSGGQSRSLAALSILAILLFK--PAPIYIL-----DEVDAALDL 1121
Qy      854 SOTIGEGQ 861
Db      1122 SHTONIGQ 1129

RESULT 80
D89824
hypothetical protein adrD [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: D89824
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D89824
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1385 <KUR>
A:Cross-references: UNIPROT:Q99W47; GB:BA000018; PID:gl3700454; PIDN:BA841751.1; GSPDB:G
A:Experimental source: strain N315
C:Genetics:
A:Gene: adrD

Query Match      3.4%; Score 172; DB 2; Length 1385;
Best Local Similarity 18.6%; Pred. No. 3.6;
Matches 210; Conservative 150; Mismatches 435; Indels 332; Gaps 47;

Qy      11 TTTTTSVTHNQEVSLVKPEILKQTOQSSSISGADYAESSGSKLKI-----NETSGPV 65
Db      29 TVGTASILVGTTLIFGLGNQEAKEASTNKELNEATTSASDNQSSDKVDMQQLNQEDNTK 88

Qy      66 DDTVTDLPSD-----KRTTPEKIKONLAKGPREQELKAVTENTESEK 107
Db      99 NDNQKEMVSSQNETTSGNKSIEKESVQSTTGNKVEVSTAKSDQASPKSTNEDLNTKQ 148

Qy      108 QITSGSQLEQS-KESLSLNTKTPVPS*SNWEICDFITKGNLVLGLS---KSGVEKLSQTD-- 161
Db      149 TISNQEGLQDPLENKSNNVVOPTNEENKVKDAKTESITLVNKSDAIKSNAETLVNNSN 208

Qy      162 -----HLVLPQSAADGTQLIQVAFPTP---DKKTAIAEYTSRAGENCEISQLVDYG 211
Db      209 SNNENNADIILPKSTAPKSLNTRMRMAAIQPNSTDSKNVNDLITS---NTTLTVVDADN 264

Qy      212 KEIINEGEVFNFSYLLKKVTIPGYKHIGQDAFV-----DNKRIAEVNLPE 256
Db      265 SKTIYPAQ---DYLKLSQITVDDKVKSGDYFTIKYSDTVQVYGLNPEDIKNIGDIKPN 321

Qy      257 SLETISDVAFALHALKQIDLPNLKAIGELAPFDNQITGLSLPQLMLRAEAPKSNHI 316
Db      322 NGETI-----ATAKHDTANNLITYFTFDVDRPNSVKMGINSIYNADATIPVDK-- 371

Qy      317 KTIETFR---GNSLKV-----IGBASFDQNDLSQMLPDGLEKTESBAFTGNPGDDHNNRV 369
Db      372 KDVPFSVTIGNQITTTTADITYPAYKEADNNSI-----GSAFTETVSHGVNVEDPGYYNQV 427
```

```

Qy      370 VLWTKSGKNPSGLATENTYVNPDKSLMQESPEIDYTKWLBEDFTYQKNSVTFGSNK---G 426
Db      428 V-----YVNP-----MDKDLKGAALKVBEAYHPKYPTN 454

Qy      427 LQKYQRN-KNLEIPKQNGVTTITE-----IGDNAFRNVDFQNK 464
Db      455 IGQINQNVNIIKIVRVPBGYTLNKGVDVNTDLVDVTFEPKNKMTYGSNQSVNLDPGDIT 514

Qy      465 L-----RKYDLSEVKLPSTIRKIGAPAFOSNNLKSFEASDDLEELKEGAPMNRRIETL 517
Db      515 SAYVMVNTKPYQYTNSESPTLVQMA---TLSSSTGNKSVSTGNAL-----GFTNNQSG-- 563

Qy      518 ELKDKLVITIGDAAPHI-NHIIYAVLPESVQVQIGRSAPFRONGANNL---IPMGSKVKTIGE 573
Db      564 -----GAGEVYKIGNYVWEDTNKNGVQELG---EKGVGNVTVTVDNNTNTKVE 611

Qy      574 MAFLSNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKKNHLKQLEVA 633
Db      612 A-----VTKEGDSYLIPLNLPNGDYRVFBSNLP---KGYEVTSPSKQGNNEELDSN 657

Qy      634 SALSHIAFNALDDNDGDEQFQNVVVKTHNSYALADGEHFIVDPPDKLSSTIVDLKILK 693
Db      658 GLSSVITVNGKDNLSAD-----LGIYKPKYNLGD---YVWEDTNKNGIQDQDE--K 703

Qy      694 LIEGLDYSTLRQTOTQPRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIA 753
Db      704 GISGVTV-TLKDNGNVKVTVTTDAG-----KYKF-----TDLNNGNY 741

Qy      754 KAERKAL---VTKCATKQGLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLVEGKG 810
Db      742 KVEFTTPEGYTPPTVTSGSDIEKDSNGLT-----TTGVINGAD 779

Qy      811 PLAQATMVQGVYLLKTPLPPEYIYGLNVYFDKSGKLIYALDMSDTTIGEGQKDAYGNPIL 870
Db      780 NM---TLDSGFY--KT---PKYNLGNVW-----EDTNKQKQDSTEKGIS 817

Qy      871 NVDEDNESGYHALAVATLADYEGLDIKITLNSKLSQTSIRQVPTAAVHRAGIFQAIONAA 930
Db      818 GV-----TVTLKNENGEVLQTKTDK-----DGKYQ-----FTGLENGT 851

Qy      931 AEAQQLLPKPGTHSEKSSSSSANSKDRGLQSN-----PKTNRGRHSAILPRTGSK 981
Db      952 YKVEFETPSGYTPQVSGT-----DEGIDSNSTTGVIKDKDNDTIDSGFYKPTVNL 905

Qy      982 GSFYV-----GILGYT-----SVALLSLITAKKKKY 1008
Db      906 GDYVWEDTNKNGVQDKDEKIGISGVTVTLKDENDKVLKTVTTDENGKY 952

RESULT 81
F82885
hypothetical protein UU482 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82885
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A:Reference number: A82870
A:Accession: F82885
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4688 <GLA>
A:Cross-references: GB:AB002145; GB:AF222894; NID:G6899476; PIDN:AAF30894.1; GSPDB:GN0001
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: UU482
A:Genetic code: SGC3

Query Match      3.4%; Score 172; DB 2; Length 4688;
Best Local Similarity 19.4%; Pred. No. 22;
Matches 217; Conservative 173; Mismatches 404; Indels 324; Gaps 55;
```



```
Db 1346 NKIAKAAAAKAEIASNP-N-LTDAEKTFTDAVDAE-VAKANDAIISAATSPADVQKEEDA 1403
QY 711 -----PRDMTAKALLSKNSLRQSGKQKPLQEAOFFLGRVOLD 749
Db 1404 GVAAIAEDVLDAAQDAKNIAKADAAAKEATGSPNLTDAEKTFTDA-----VD 1454
QY 750 KAIKAEKALVTKATKNGCOLLERSINKAVLAYNNSAIKANKVKELEKELDLTGLVSGK 809
Db 1455 AEVAKANDAIISAATSPADVQKEEDAGVAAIAEDVLDAAQDAKNIAKESDAKSAIDAN 1514
QY 810 GPLAQTWVQGYLLKTLPLPEYYIGLVNVPDKSKGLIYALDMSDTIGEGOKDAYGNPI 869
Db 1515 PNLTDAB-----KESAK-KAVDADAKAATAIDASTSPV 1547
QY 870 LNVEDNNGYHALAVATLADYBGLDKITLNSKLSQTSIRQVPTAAVYHRAGIFOAIONA 929
Db 1548 EAQSAEDKGVGSIQDVL-DAAKQDAK---NKIAKEVAAAKEAIDA---NPNLSDAEKEA 1600
QY 930 AAEAEQLLPKPGTHSEKSSS---ESANSKDRGLQS 962
Db 1601 SKKAVDADAKATDAIDASTSPVQAEDKGVGS 1635

RESULT 83
D97316
Probable S-layer protein [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: D97316
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: D97316
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1939 <KUR>
A:Cross-references: UNIPROT:Q97DT3; GB:AB001437; PIDN:AAK81319.1; PID:g15026474; GSPDB:G
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3389

Query Match 3.4%; Score 171.5; DB 2; Length 1939;
Best Local Similarity 19.1%; Pred. No. 6.3;
Matches 201; Conservative 151; Mismatches 386; Indels 315; Gaps 49;

QY 12 LTTVSVV-----THNQEVSLVKEPILKQTOQASSISGADYAESGSKLKINETSG-P 64
Db 907 LPTINVTGNSTQWTKNDVVLNVSKT-----AGASGVSKVTVIKNDGTP 949
QY 65 VDDTVTDLFS-----DKRTTPPEKINDILAKGPR-----EDEL 96
Db 950 VDIENSSYTVGNGNYIPTVINGAGASASTVVVVDK---IDKVPKSVAVSSMKIGEDY 1006
QY 97 -----KATENTSEKQITSGSQLESKESLSLK-----TVPTSNWEICDPIIT 141
Db 1007 NYSATVPAEKDVTVALQNPNNVASGVKYYISTNGTDMNPIDEGTSYTVTSTGN-----AT 1060
QY 142 KGNLTVLGSKSGVEK--LSQDHLVLPQAADGTOLIQVASFAPTPD----- 186
Db 1061 VNYKFAVNSGSGVEDVQKQVYIIPKPYT---NNIVTATTSNGADYNGQWTKNDINPV 1117
QY 187 -----KKTAAIAYTGRAGENGHISQLDVG---KEIINEGEVFNISYLLKKVYIPTGYKH 238
Db 1118 LSGGISAEAFVQYQSTDPDSDEASWKMCGANFNELTVSKQDAKYFRAVSKGSGSV 1177
QY 239 GQDAFVDKNNKIAEVNLPSPLETISYAFAPHAALKQIDIDPNLKAIGELAFF----- 289
Db 1178 TSG-----AELKLDKTIPTAMT-----VALKENKFTFPLTISPNLFFKOTLDVNIY 1224
QY 290 -DNQITG-----KLSLPRQLMRLAERAFKSNHIKTIET 321
```

```
Db 1225 ANNDISGIDHYDYOVEDGTGDAYSENGPMKTYSELKLSRQKGV-----YARADK 1277
QY 322 RGNLSKVIAGEASF-QDNDLSQMLPDGLEKTESBAFTGNPGDDHNNRVVLTWTKSGKPNPS 380
Db 1278 AGNISKVVSSDGFIVDNN-----KATAPKITATAGDKTYNS--WTASDIN-- 1321
QY 381 GLATENT-----YVNPDKSLWQESPIDYTKWLEEDFTYQKSVTVTFSGKGLQK- 429
Db 1322 -IVLSNSTAFSDIAGYQYKAGSNGTWTDMP--LSSDGLTKDKKITINQNTKNKLYYFRAVSKS 1379
QY 430 -----VKR-----NKNLEIPKQHGVT-----ITEIGNAFRNVDFQNKTLRK 467
Db 1380 GYCGEENSLMIRRDALAPSTKVDTGVANSWTSKSVFVSNANDNIIAPISYIK-IGS 1438
QY 468 YDLBEVKLPSTIRKIGAPAFQSNNLKSFASDD--LBEIKEGAFMNNRIETLELKKDLVT 525
Db 1439 GDMN--KMDSTY--TFNNEINSAGVQKAVSDAGVENV-----SNTIYNVMIDKTTPT 1487
QY 526 IGDAA-----PHINHIYAIVLPESVQEIQRSAFRQ-NGANNLI FMGSKVKTILGENAFLSNR 580
Db 1488 ITGAAQNGSYSIGRV--IGYNDNFQIETATYKKNNGAAQNLTSQSOLTTPFGAYS----- 1540
QY 581 LEHLDLSEQKOLTEIPVQAFSDNALKEVLLPASLK---TIREBAFKKHHLKQLEVASAL 636
Db 1541 LEVVDKSGNRSVLNFTLKALP--SVNDVLYTSDSKNAIDAIRNELNTHNDLKE-PYASQI 1597
QY 637 SHIAPNALDDNGDQFDNKVVKTHNSYALADGEHFI VDPDKLSSTIVDLEKILKLIIE 696
Db 1598 -----KASVKALEDKRYAALEKE--VMSLKDOTSNIV--KTVDLSLPN 1634
QY 697 GLDYSLRLQTTQTOFRDMWTAGKALLSKNSLRQEKQKFLQEAQFFLGRVDLKA----I 752
Db 1635 GKOGIIALENPKMSVYDKIAGG-----SSTLTNSQYAKLINESYLDKDKLNAIKALKDQV 1689
QY 753 AKAEKALVTKATKNGCOLLERSINKAVLAYNNSAIKANKVRLKELDLTLGLVEGKGPL 812
Db 1690 ARVEAAI-----KALPA--KGSVKKADSVSKIKAVNDSYNALNKEQKAL 1730
QY 813 AQATWQGVYLLKTLPLPEYVIGLVNVPDKSKGLIYALDMSDTIGEGOKDAYGNP--IL 870
Db 1731 VNPDLVK-----KLNDVVDKASKLLHLHNSPNGITVTGVDGTFKFSPTVL 1774
QY 871 NVDEDEGYHALAVATLADYEGLDIKTILNSKL 903
Db 1775 VVDPIKEGNGQTEFNSSGADY--VKKATLTNDKL 1805

RESULT 84
T30336
nuclear/mitotic apparatus protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30336
R:Merdes, A.; Ramyar, K.; Vechio, J.D.; Cleveland, D.W.
Cell 87, 447-458, 1996
A:Title: A complex of NuMA and cytoplasmic dynein is essential for mitotic spindle assem
A:Reference number: Z20828; MUID:97053784; PMID:8898198
A:Accession: T30336
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2253 <HER>
A:Cross-references: UNIPROT:P70012; EMBL:Y07624; NID:g1514670; PIDN:CAA68905.1; PID:g151
C:Genetics:
A:Gene: NuMA

Query Match 3.4%; Score 171.5; DB 2; Length 2253;
Best Local Similarity 18.8%; Pred. No. 7.9;
Matches 226; Conservative 188; Mismatches 426; Indels 363; Gaps 53;

QY 6 KTVALTITVSVVTHNQEVFSLVKEPI--LKQTOASSISGADYAESG-----K 53
Db 125 KQTQELASILRFVLDNEDALCLDDKLIINFLKR-KAPFPSSGNDPSSSSDAMTPNVSHRRK 183
```





QY 620 EAFKNHKLQLEVASALSHIAPNALDDNDGDEQPDNKVVVKTTHNS-----YAL-----AD 670  
Db 619 VCTPKYEIQNEEYIPLNGI-PS-----DNEIYTEQKSLDKIEYPLTGVFSD 666  
QY 671 GEHFTVDP-DKLSSTIVDLKILKIEGLDYSTLRQTOTQFRDMTTAGKALLSKSLNRQ 729  
Db 667 NETYITEHIGSLNEICPLNEVL-----LDNEIITEQIGSHDEEYIPLNEILPNEVIT- 720  
QY 730 GEKQFLQEAFFLGRVLDKAIAKAEKALVTAKATNGQLLER-----SINKAVLAYN-- 783  
Db 721 -EQDGLDEEYIPLTGVFSDN-----ETVITEQKSLDKIEYPLTGVFSDNETYITEHIG 774  
QY 784 --NSAIKXANVKRLEKEL--DLITGLVKGKPLAQATWVQGVYLLK-----TPLPLPEY 833  
Db 775 SLNEEICPLNEVLNDEIITEIGSLDEAICPLNEILNDEISTLKEKDSQNEICTPKQ 834  
QY 834 YIGLVVYFDKSGKLT--YALDMSDTIGEGQDAYGNPILNVDNDEGHALAVATLAD-- 889  
Db 835 EVSQNEEINTLNEILSDNEIHITEQIGSPNEEY--PLTGVSSDNETYITEHIGSLNEEI 892  
QY 890 -----YEGDITILMSKLSQLSIRQVPTAAVHRAGIFQAIQNAABAEQLLPKPGT 942  
Db 893 CPLNEVLNDEIITEQIGSNNGMYSLNEV-----PSDNEISTPKQKDSQEBEISTPK--- 945  
QY 943 HSEKSSSESANSKDRGLOSPKTNRGRHSAILPTGSGKSPVVGILGVS 993  
Db 946 --QKGSQEEINILNEILSDN-----EIHITEQIGSPNEEYIPLTGVFS 987

RESULT 86  
JH0720  
tanabin - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 10-Jun-1993 #sequence\_revision 10-Jun-1993 #text\_change 09-Jul-2004  
C:Accession: JH0720  
R:Hemmati-Briavanlou, A.; Mann, R.W.; Harland, R.M.  
Neuron 9, 417-428, 1992  
A:Title: A protein expressed in the growth cones of embryonic vertebrate neurons defines  
A:Reference number: JH0720; MUID:92398961; PMID:1524825  
A:Accession: JH0720  
A:Molecule type: mRNA  
A:Residues: 1-1744 <HEM>  
A:Cross-references: UNIPROT:Q01550; GB:M99387; NID:g214816; PID:AAA49966.1; PID:g214817  
A:Experimental source: tadpole head  
C:Keywords: intermediate filament

Query Match 3.4%; Score 171; DB 2; Length 1744;  
Best Local Similarity 17.7%; Pred. No. 5.7;  
Matches 204; Conservative 203; Mismatches 451; Indels 296; Gaps 50;

QY 17 VYTHNQEVFSLVKE--PILKQ--TQASSISGAOYAESGSKLKNINETSQVDDTVTDL 72  
Db 656 IISKRKVF-LENEYIPVSKDOLTFTSHLE--NDSESSQSPDSKLFENKSTEDQLITNL 712  
QY 73 FSDKRTTPEKT-KNLAKPREQLKAVTENTESEKQITSGSQLESLSLNKTVPEST 131  
Db 713 ---KSTQENIQSN-----QEHLENLEFDSVPEDT 740

QY 132 SNWEICDPTTKGNTLVGLSKSGVEKLSQ--TDHLVLPQAAAGTQLIQVASFAPDPKKT 189  
Db 741 VKFM---YPQENNLLEENVYGDGLVQWATDENII-NQSSDQLLSDHSHHEETKTS 796

QY 190 AIAEYTSRAGENGEEI---SQLDVDGKEINEGEVFNYSLLKKVTIPTGYKHIGQDAFVN 246  
Db 797 IAVEHNRMESEHAEDVKSSEIPVEISENVSEVEIIEH--ISDVEEDT-----KQAFED 848

QY 247 KNIAEVLNPLESLETISDYAPAHALKQIDLPNLKAIGELAFPNQITGKLSLPQLMR- 305  
Db 849 RVGEIQINQNEESTVDLGSVYSQENSQLEDEVSISEQEKOFINEQCLSKSDQIRE 908

QY 306 -----LAERAFKS--NHIKTIEPRGNLSKLVIGASQDND-----LSQLMLP 345  
Db 909 APTDEEDVHVVDVFMQEQSFREVQQLNNIKQEDVDLYNQYDEDSFQNNDEBQELSCDLQ 968

QY 346 DGLEKIESE-AFTGNPGDDHY-NNRVVLWTKSGKQNPGLATE-----NTYVNP 391  
Db 969 EQKIKLEENQLSNEGQNFQGGNDIEFSGQYDTTDEICQETIGNQVSAQLLCESDINQ 1028  
QY 392 DKSLW-----QESPEIDYTKWLEBDFTYQ---KNSVTGFSNKGKQVKRNKNLBIKPKOHN 443  
Db 1029 DKLMEDEBEQNNPETEDNIGLESDQENSTRNEGTFQSEBCDVTVPKEDMDKSEYS 1088  
QY 444 G-----VTITE-----IGDNAPR 456  
Db 1089 GQOEDLQKQVTPSLNEQANNDLLEKEBVLHADDQORSVNDITIDKLSERIIDNELA 1148  
QY 457 NVDQNTKL-----RKYL--BEVKLPSTIRKIGAFQSNILKSFASDDL----- 501  
Db 1149 TVDV-NESLAANKQVDFLFTDEYAVDDNV-----GMQDDSGQYQTKEDLFVDGNIIIE 1201  
QY 502 -EEIKGAGFNNRI-ETLELKDCLVTIGDA---AFHINHIYAIVLPSVQEIGRSAPRON 556  
Db 1202 KIEIQOISLLNQEICERVDNVEDIS-GEAKNESVEMNDVVDLPREAKVTGDEQISPLQD 1260  
QY 557 GANNLIIPMGSKVKTGEMAF-LSNRLBHLDLSEQQLT-----EIPVQ-----AF 600  
Db 1261 EKLNETMEDTKDNDGQCLCEKENETEYIEVTDSPQFATDLSDHAGRELTVDQNSANLQF 1320  
QY 601 SDNALK-----EVLPLPASLKTIRBEAPKQKHLKQLEVASALSHIAPNALDDNDGDE 651  
Db 1321 CENPTKTLIAHHIETVADSDLESTEEQVQETERIIPFKPEDSKHEN-----ENSESEE 1374  
QY 652 QFDNKVVVKTTHNSVALADGSHFTVDPD-KLSSSTIVLEKILKIEGLDYSTLRQTOTQ 710  
Db 1375 SVDSQBSILSHKS-----EEPEISKDYQLEQLPDTVPLPNL-----EDE 1415  
QY 711 FRDMTTAGKALLSKSLNRQSEKQKFLQSAQPFGLGRVDLDKAIAKAEKALVTAKATNGQL 770  
Db 1416 FEDLT-----EQPDVHEEHQNDGSCASTFITSVDEDK-----EREVRESVSKDEES 1462  
QY 771 LERSINKAVLAYNNSAIKXANVKRLEKELDLTLGLVKGKPLAQATWVQ-----VYLL 824  
Db 1463 NEEFQGVVLSVDKTSQVEVITLGLAQSPSYLGNBESSEDSMENARILNPNFNDIVDFM 1522  
QY 825 KTLPLPEYIIGLVN-----YFDKSGKLIIVALDMSDT-----IGSGQ 861  
Db 1523 VSQMTETKIIIAEQVTEVTEVTFQDPAKNTLNAREKETDYDSNEENIEFTNEQ 1582  
QY 862 KDAYGNPIL-----NVYDNEGYHALAVATLADYEGE-----DIKTLNLSKLSQTSIR 910  
Db 1583 SASPANDIVDENQSEDSVISDNQ-----TTSSEYEDLPNATSIHVVALEESNISITTE 1635  
QY 911 QVPT-----AAYHRAGI--FQAIQNAABAEQQLPKPGTHSEKSSSSSANSKDRGLQ--- 961  
Db 1636 QSSTDTRKMTYEGYEITSLQNVEDNAQETEKEPFSPVPLQGEDRSRDEELDDEGSEFSF 1695  
QY 962 -SNPKTNRGRHSI 974  
Db 1696 GVNDEKANGHKDV 1709

## RESULT 87

D96796

probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C:Accession: D96796

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dwyer, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

Qy	30	EPILKQTOASSISGADYVASSSGSKULKINETSQPVDDTVT-----DLPSDKRTTEK 82
Db	535	QPAEESTTNSEKVPDTSSENTEG-----VSSNPSDTTSVGBSNKPEHNDKSNENSEK 588
Qy	83	IKONLAKGPREQELKAVTENTSESKQITSGSQSQSKESLSLNTK-----VP 129
Db	589	TVBEPVNPNEGTVEG-TSQNETSKPQPAEETQNSKIANENTGEVSNKPSDKPPVE 647
Qy	130	STSNWEICDFITK-----GNT-----LVGLSKSGVEKLSQTDHLVLPQQAADG- 172
Db	648	ESNQPEKNGTATKPSNGSGNTTSENGQTEPEKKLELRNVSQDIELYSQTNGTQRQHVSLDGI 707
Qy	173	-----TOLIQVASPAF-----TPDKK-----TAIAEYTSRAGENGELSQ-----206
Db	708	PENTDTYFVKVKSFAKDVIPIVASITEERKNGOSVYKITAKAEKLOELENKYVNDFTF 767
Qy	207	LDVDGKEIINEGVSFNYS--LLKKVT- IPTGYKHIG-----QDAFV 244
Db	768	YLDCKAKE-----ENTNFTSFNLVKAINQNFSGTYHLAASLVANAEVLGPDERSYKIDFT 824
Qy	245	-----DNKNIAEVNLPESL-ETISDYAFAHALKQIDLDPNLKAIGELAFPDNQITGK 296
Db	825	GRLTGEXDGNRYATYNLKPFLFENLSGATVEKLSKNV-----AISGK 867
Qy	297	LSLPROLMRLAERAFKSNHIKTITFGRNSLKVIGBASFQDNDLSQLMDGLGKETESAF 356
Db	868	-----NDIGSLANEATNGTKIKQHVHDG-----VLJGER-----GVGGLAKAD 906
Qy	357	TGNPGDDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQK 416
Db	907	QSSIAEISSFKGRIV-----NTYETD-----927

QY 417 NSVTGFSNGLQVKCNKLNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKVDLEEVKLP 476  
Db 928 ----AYNTGGVLGHJTGKNASTAKSATVTTIS----SNTNRS-----D 962  
QY 477 STIRKIGAFAPQSNLNKSFASDDLEEK-----GAPMNRRIE-----TLELK 520  
Db 963 QTVGGLAGLVQDQDAHIQNSYAEGLDINNKKHGVKAGVAGYLVDRTSBEKGIAGELTNVL 1022  
QY 521 DKLVTIGDA--AFH-----INHIIYAVLPESVQIEGRSAFRONG-----ANNL 561  
Db 1023 DVNVTVNGNAITGYHTGMKVANTPSSKANRVFNVTL-EKDEWVSKESPEERGTMLDASQI 1081  
QY 562 IPMSKVTIGEMAFNLSNLEHLDSLSEQQLTEIPVQAFSDN-ALK-----EVLIPASLKT 616  
Db 1082 VSKKABINPL-----TLPTVPLSTSGGKSDFSKIAHYQANRALVYKNIIEKLLPFFNYS 1136  
QY 617 -----IREEAPKKNHLKOLEVASALSHIAFNALDDNDGDEQPDNKKVVKTHNSYA- 667  
Db 1137 TIVKGNLVKENS-----LYQKELLSAVMMKDDQVITDIVSNKQTANKLLI--HYNDHSS 1190  
QY 668 -----LADGEHFIVDPDKLSSTIVDLEKILKILIEGLDYST- 702  
Db 1191 EKFLQKYQTFANLAENYLNGLTGLYTPNQFLYDRDSI-----VKEVLPLOKLDYQSD 1244  
QY 703 -LRQT-----TOTQFRDWTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLXAI 752  
Db 1245 AIRKTLGISPEVKLTLEYLEQ-----FSKTKQNLGDSLKLLSADAGLAS---DNSV 1294  
QY 753 AK--ABKALVTKKATKG-QLLERSIN-----KAVLAYNNSAIKKANVKLEKELD 800  
Db 1295 TRGYLVDKIKKNEKALLGLYLERKWNFTNGQVNVKDLVMHDPFGKNGTSPDLTILIE 1354  
QY 801 LLTGLVEGKPLAQATMVQGVYLLKTLPLPEYIYGLNVYFDKSGKLIYALDMSDTTGE- 859  
Db 1355 L-----GKS-----GFNNLLAKNNVDTYGISLASQHGAT 1383  
QY 860 ---GQKDAYGNPILNVDEDNREGYHALAVATLADYEGL--DKITILNSKLSQLTIRQVPT 914  
Db 1384 DLFSTLEHYRKVFLPNTSNDFKXSETKAYIVEKSTTIEVKT-----KQGLA 1431  
QY 915 AAYHRAGIFFOAINAAABEOL-----LPKPGTHSEKSSSESANSDKRGLOQNPKNR 968  
Db 1432 GTKSIGVYDRITSATWKYRNWVPLLTLPERSVFVISTMSLGFAGYDRYRSSDHKAGK 1491

RESULT 89  
A:26655  
A:myosin heavy chain [similarity] - elime mold (Dictyostelium discoideum)  
N:Contains: myosin ATPase (EC 3.6.4.1)  
C:Species: Dictyostelium discoideum  
C>Date: 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 09-Jul-2004  
C:Accession: A26655; A24728; S00250  
R:Watrick, H.M.; De Lozanne, A.; Levinwand, L.A.; Spudich, J.A.  
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986  
A:Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium disco  
A:Reference number: A26655; MUID:87092266; PMID:3540939  
A:Accession: A26655  
A:Molecule type: DNA  
A:Residues: 1-2116 <WAR>  
A:Cross-references: UNIPROT:P08799; GB:M14628; GB:M11938; NID:G167834; PIDN:AAA33227.1;  
R:DeLozanne, A.; Lewin, M.; Spudich, J.A.; Levinwand, L.A.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985  
A:Reference number: A24728; MUID:86016788; PMID:3901008  
A:Accession: A24728  
A:Molecule type: mRNA  
A:Residues: 2035-2116 <DEL>  
R:Wegle, G.; Noegel, A.; Scheel, J.; Gerisch, G.  
FEBS Lett. 227, 71-75, 1988  
A:Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium  
A:Reference number: S00250; MUID:88112226; PMID:2828113  
A:Accession: S00250  
A>Status: nucleic acid sequence not shown

A:Molecule type: DNA  
A:Residues: 1734-1893 <WAG>  
C:Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phos;  
F:1-818/Domain: globular head <HED>  
F:89-747/Domain: myosin motor domain homology <MMOT>  
F:179-186/Region: nucleotide-binding motif A (P-loop)  
F:819-2116/Domain: alpha-helical rod <ROD>

Query Match 3.4%; Score 171; DB 2; Length 2116;  
Best Local Similarity 20.0%; Pred. No. 7.6;  
Matches 213; Conservative 163; Mismatches 373; Indels 316; Gaps 51;  
QY 55 KLIKNETSGPVD--TVTDLFSDKRTTPKIKDNLAGPREQL-----KAVTENT 103  
Db 897 EIRVEDMESELDEKKALENLQNKRSVVEKVRDLBELQEQKLRNLTLEKIKKKEEEL 956  
QY 104 ESEKQITSG-----SOLEQSKESLSLNKTPSTSNWEICDPTIKGNTLVGLSKSGVEKLS 158  
Db 957 EEMKRVNDGSDTISRLEKIND--ELQKEVEELTE-----SFSEESKDKGVLEKTRVRLQS 1010  
QY 159 QTDHLV--LPSQAADGTOLI-----QVASFAFTPDKTAIAEYTSRAGENGESIS 205  
Db 1011 ELDDLTVRLDSETKDKSELRLQKKLEELKQVQE-ALAAETAAKLAQEAANKKLQGEYT 1069  
QY 206 QLOVDGKEIINEGEVFNYSLLKKVTIPTGYKHI-GQDAFVNDKNIAENVLPESLETISDY 264  
Db 1070 ELN-----EKFNSEVTARSNVESKKTLESQLVAVNNELDEEKNRDALEKKKKA 1119  
QY 265 APAHLALQIDLPNLKAIG-----ELAFDNOITG-----KLS 298  
Db 1120 LDAMLE-----EMKQLESSTGEEKSKSLYDLVKVQESDMEALRNQISELQSTIAKLEKIKST 1175  
QY 299 LPROLML-----AERAFKSNHIK-----TIEPRGNSLKVIGE-ASFQDNDLSQLMLPDG 347  
Db 1176 LESEVALQSELEAEQAKSNVEKQKKVLDLEDKSAQLAEETAAKALDKLKKLEQE 1235  
QY 348 LEKIE--SEAFPTGNPGDDHYN-----NRVLMVTKS-----GNPSPGLATENTYV 389  
Db 1236 LSEVQTQLSEANNKNVNSDSTNKHLETSPNNLLELEBAEQAKQALEKCRGLGSELKHV 1295  
QY 390 N---PDKSLWQESPE---IDYTKWLE-----EDPTYQKNSVTGFSNK---GLQKVK--- 431  
Db 1296 NEQLREBKQKESKSKKRVLEKEVSELKQDIEEVASKKAVTEAKKQKSELDEIKRQY 1355  
QY 432 -----RNKNLEIPK--QHNGVTITETGDNAFRNVDFQNKTLRK--YDLSE----- 472  
Db 1356 ADVVSSRDKSVEQLKTLQAKNEELRNLTAEAEQGLDRAERSKKKAEFDLEAVKNLEET 1415  
QY 473 ---VKLPSTIRK-----IGAPAFQSNLNKS--FEASDD--- 500  
Db 1416 AKKVAEKAMKKAETDYRSTKSELDDAKNVSSEYQVQIKRLNEBELSELRSVLSEADERCN 1475  
QY 501 -----LEEIEGAFPMN-----RIETLE--LKQKLVIG----- 527  
Db 1476 SAIKAKKTABSALSELADEIDAANNAKAKAERKSKELEVRVAELESLEDKSGTGVNVEFI 1535  
QY 528 ---DAAPHINHIVAVLPESVQIEGRSAFRONGANNLIFMGSKVKTGLGEMAFNLSNLEH- 583  
Db 1536 RKQDA--EIDDLRALRDRETSRIKSDKDKKNTKQFADLEAKVEEAQREVVVTDRLKK 1593  
QY 584 ----LDLSEQ---KOLTEIPVQAFSDNALKEVLL-----PASLKTIRREAFKK--NHL 627  
Db 1594 LESDIIIDLSTQDQDTEKRIKIEK-SKKLEQLTLAERRAAEBEGSSKAADBEIRKQVQEV 1652  
QY 628 KQLEVASALSHIAFNALDD-----NDGDEQPDNKKVVKTHNSVALADGEHFIVDP 678  
Db 1653 DELRAQDLDSRAALNASEKKIKSLIAEVDEYKQLEDEILAK----- 1694  
QY 679 DKL-----SSTIVDLEKILKIE-----GLDYSTLRQTTQTQ----- 710  
Db 1695 DKLKAKRALEVELEEVLDQLEEEEDSRSELEDSKRLTTEVEDIKKKYDAEVQNTKLD 1754

QY 711 -----FRDMTTAGKALL-SKSNLRQGE--KOKFLQEAQFFLGRVDLD-KAIKAKAEKALV 760  
Db 1755 EAKKKLTDDVDTLKKQLEDEKKKLNESERAKKLESENEDEFLAKLDAEVKNESRAE---- 1810  
QY 761 TKKATNGQLLERSINKAVLAYNNSAIKANVYK----RLEKELDLLTGLVEGKGLAQAT 816  
Db 1811 -----KORKKYBKLDKDTKYKLNDEAATKTQTEIGAALKEDQIDELRSKLEGE--QAKAT 1863  
QY 817 MVQGVLLKTLPLPEYIYGLNVPYFDKSKLIVALDMSDTIGEGQ 861  
Db 1864 QAD-----KSKTLEGEINLRAQIEDEKIKWRLEKEKRALEGE 1903  
RESULT 90  
G88545  
protein F59B2.12 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustli.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: G88545  
A:Status: preliminary  
A:Molecule type: DNA  
A:Cross-references: UNIPROT:F34487; GB:chr\_III; PIDN:CAA77581.1; PID:G3877841; GSPDB:GNO  
A:Residues: 1-918 <STO>  
C:Genetics:  
A:Gene: F59B2.12  
A:Map position: 3

Query Match 3.4%; Score 170.5; DB 2; Length 918;  
Best Local Similarity 18.6%; Pred. No. 2.3;  
Matches 176; Conservative 142; Mismatches 359; Indels 267; Gaps 41;  
QY 25 FSLVKEPILKQTAQSSISGADYAESGSKLKNETSPVDVTVDLFSKRTTPEKIK 84  
Db 55 FAMPKLDASKAAVHSSSHKHGHSSGSS-----SNTHSLTVVGADGKNITENSEK 107  
QY 85 DNLAKPQREQLKAVTENTESEKQITSGSQLESLSLNTKTPSTNWEICDFTKGN 144  
Db 108 DGYNK---ESKVDENANTKIKS--ADGSVITGK---SHNKSDDASY----- 149  
QY 145 TLVLGSKGVEKLSQTDHLVLPQADGTQLIQVAFAPDKKTAIAEYTSRAGENG 204  
Db 150 -----GLEKSSKT-----YADKNGTMLSNTNKINNQSR 179  
QY 205 SOLVDVGKEI NEGEVFNLSYLLKVTIPTGYKH-----ICQDAFVDNK 247  
Db 180 AALD-EGNEFVNQADGTFLRN-----TGHKNTDEHLNHLVDNAQMSIGADGTSNI 234  
QY 248 NIAEVLNPLESETISDYAFALHAKQIDLPDLNK-----AIGELAFPDNOI---- 293  
Db 235 TWRKSGVGDHNAAD--AHNPFESLDAGNKKQNTYKKAASAGSNADFNESLSK 291  
QY 294 -----TGKLSLPRQLMPLAERAFKSNHIKT-----IEPRGNSLKVIGBASTQD 336  
Db 292 NADGTSMSNSTGNFNNTSYDKATAEVMGSKVNVADGTSSMEASHAGSNSSKINSASQS 351  
QY 337 NDLSQLMLPDGLEKLESEAFNPGDDHYNNRV-VLWTKSGKNPGLATENTYVNDKSL 395  
Db 352 SDSL-MVGNIGTKSHSTNKTNDYALDEANQASGSI SEQIGKNGQORSLNESSIESGRKAE 410  
QY 396 WQESPEIDYTKVLEBDFTYQKNSVTGFSNKGLOKVRKNLRI PKOHN-GVTITIGNA 454  
Db 411 SRNTAADTLDSVDANGTVSSSHSKSASGTSLSDE--NHN-----KTHALQASVDEHGNK 463  
QY 455 FRNVD--FQNKTLRKYDLEEVKLPSTIRKIGAFQSN-----NLKSFASDDL-----E 502

Db 464 NHSIDGYSRNNKKTGTFGNSE--MSASIKNADGTMQVNVKNDTNRNTYEAEKSALEKONHE 521  
QY 503 EIKEGAF-----MNRRI-----ETLELKDKLVTIGDAAPH 532  
Db 522 KNSDGTFOEKGSGNSRNVRTDGGSNLAVGYSVGKGGVSSNETTASSNFAFTSADSNQ 581  
QY 533 INHIY-----AIVLPESVOEIGRSAPFRQANNLI-----FMGSKV-KTLGEMA----- 575  
Db 582 FDHLHOKTANGTEITHAKDSKQVAASANAKSSLDTSMSAVDAKGNKVDKTSQAADSHDA 641  
QY 576 -----FLSN-----RLEHLDLSEQK-----LTEIPVQAF 600  
Db 642 ISASDVEDAKIVKHADRSIESISNDSSNQTSASEHNSDKQSEHEKRNQADGSPFSSNSA 701  
QY 601 SDNALKEVLLPASLTKTIREAFKK-----NHLKOLEVASALSHIAFNALDNDNDGEQPN 655  
Db 702 KNAVKDA-----TDVQNAPFANVDAAGNSVSEKNSLVENH-----SNDSDASSD 748  
QY 656 KVVVTHNSVALADGEHFI VDDPKLS-STIVDLKILKLIIEG---LDYSTLRQTQTQF 711  
Db 749 KIHKSVDGT-ELTDKHSNVSTSHLTGQLAQIQKLMADGVSVMNSDNSHITHEKSR 807  
QY 712 RDMITAGKALLSK-----SNLROGEKQFLQEAQFFLGRVDLDKAIKAEKALVTKATK 766  
Db 808 ADVDAKHKAHLKADGSGSTIDIEG-----FKHSDLESRGEGAQK----QRYQK 852  
QY 767 NGQLLERSINKAVLAYNNSAIKANVNRKLEKELDLTLGLVEGK 810  
Db 853 LONGTESND---VGYEKMAKGGDTSSHKK---TLAKDGGK 889

## RESULT 91

S61535

nucleotide-binding head-stalk protein 183K - Giardia lamblia

C:Species: Giardia lamblia

C:Date: 23-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 09-Jul-2004

C:Accession: S61535

R:Marshall, J.; Holberton, D.V.

J. Cell Sci. 108, 2683-2692, 1995

A:Title: Giardia gene predicts a 183 kDa nucleotide-binding head-stalk protein.

A:Reference number: S61535; MUID:9606038; PMID:7593309

A:Accession: S61535

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1620 &lt;MAR&gt;

A:Cross-references: UNIPROT:Q24984; EMBL:X79815; NID:9871047; PID:9871048

Query Match

Best Local Similarity 3.4%; Score 170.5; DB 2; Length 1620;

Matches 228; Conservative 184; Mismatches 426; Indels 325; Gaps 52;

QY 21 NOEVPSL-----VKEPILKQTAQSSISGADYAESGSKLKNETSPVD 68

Db 254 NOEVNLRVTQVSRDEILQDKSMLISTLQNIQVMAEKVDAAQAPFVARNDLS-----L 308

QY 69 VTDLFSKRTTPEKIKONLAK-----GPREQ-----ELKAVTENTESEK 108

Db 309 AEETIKESVTIEKLMKLSKTSQVAIEQRDODAVELARSVADMKQLSDTVEKQR 368

QY 109 -----ITSGSQLESLSLNTKTPST--SNWEICDFTKGNITLGLSKSGVEKLSQ 160

Db 369 IVHQLHIDMSKKSHEBSISQHKSEIAKSAFDESTFIKNAR---ISKLTTE-LTET 423

QY 161 DHLVLPQADGTQLIQVAFAPDKKTAIAEYTSRAGENGESIQLDVDGKEI NEGEV 220

Db 424 Q-----TQI-----ASALEKQELQSESCKKAQAHAEEV-----LLAEREI 457

QY 221 FNSVLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNPLESETISDYAFALHAKQID--LPD 278

Db 458 ANTLTQELQATK-----GELEAVRKSNEHQVNSYESLQKTQOSYEALLREKETIDKL 512

QY 279 NLKAI GELAFPDNQITGKLSLPRQLMPLAERAFKSNHIKTIFRGNLSKVI 329

Db 513 NLECDNAKAFDDHSEQSTAHYQALREEVERVRHTSENLLQEKMGVINRLTAETIOAIKLD 572  
Qy 330 GEASFQDNDLSQMLPDGLEKLESAFTGCPDCHYNNRVVLTWTKSGKPSG----- 381  
Db 573 GERALDSKDTQIAELQKNVSSLMEDL-----TKTSSNSAEISRLTNL 615  
Qy 382 -LATENTYV---NPDKSL-----WQSPEDITYKWLBEPTYQKNSVTFGSKG-LQK 429  
Db 616 LEATRKDYVQRLESQDEQLRATNEYKENWEKLDVAFPAKDAELSKQTAALHALRDELSQ 675  
Qy 430 VGRNKNLEPK-----QHNGVTITEIDNAPRVDFQNTLRKYDYDLEEVKLPSTIR 480  
Db 676 VKNHLLVEDRLKVBNSAQNTSMTL-----LSNQEISRLGRLDV--VKNSASIN 726  
Qy 481 KIGA-FAPQSNLKSFEASDDLEETKE--GAPMNRRIETLE-----LKDKLVITGDAAFH 532  
Db 727 EVEANKTKQSLNLTISTQELQDSKNEADAALKEKVELDTLRLMDSDSTAAKAQN 786  
Qy 533 INHIVAIVLPESVQBIGRSAPRQNGANNLI FMGSKVKTLGEMAPLSNRLEHLDLSEOKQ 592  
Db 787 LQN-----TVDKLQBELSSVS-----SDKLAI SAENRVVSELKH-ELSTEXQM 829  
Qy 593 TEIP-----VOAFSDNALK-----EVL-LPASLKTIRBE-----AFKQHLKQLEVAS 634  
Db 830 REAENSRAQLQISHLEAQVKDAKAKADSEVTRLLSDLKTVEKELAIIVVDQDAKIAEISR 889  
Qy 635 ALSHTAFNALDDNDOGFQDNVVKVTHNSYALADGEHFI-VDPDKLSSTIVLEK--- 690  
Db 890 KLED-TPHKLQRS--EQTVEVLQAABEKELSVAKLNTDQTIALNDRVANLAAELNKQKN 946  
Qy 691 ---ILKLEGLDYSLR---OTTOTQPRDMTTAGKALLSKNLRQGEKQKFLQEAQPLG 744  
Db 947 ETEELLAFKDYKQKQKLEDSRTEVTETSSGRABEISRLQLQIDNIGELALLQAIYA 1006  
Qy 745 -----RVDLD--KAIAKAEA-----LVTKKATKNGQL----- 770  
Db 1007 QKEDALKDOLNSAKAVLLAESAEKDAIISTLKQDLTNLRAELSSSEAKDVTIARYKQDC 1066  
Qy 771 -----LERSINKAVLAYN-----NSAIKKNVVRLEKELDLTG-LVEGKGP 811  
Db 1067 ENLOTSLTSEKKEEAYNLIKQEPAGYKQVSAKQA-----YEAQIASLTGDLAAAKK 1122  
Qy 812 LAQATMVQGVLLKTPLPPEYVIGLVYFDKSGKLIYALD-----MSD 855  
Db 1123 SEQLEMEIEREMKHASASKQAQELKVTIDQKQRELELOKEYLSSQAEALSTWRLQD 1182  
Qy 856 TIGEQKDAYG--NPILNVDEDNEGYHALAVATLADYE-----GLDIKTIL----- 899  
Db 1183 BIDKLIQDKGLINQISSLIKQE-----ISVSS-ADREHQAKKADKDKITLQDALRIAE 1235  
Qy 900 -----NSKLSQ-----LTSIRQVPTAAYHRAGIFQATON--AAABA 933  
Db 1236 DKRVLEEGNSKLEBIKQYECLISEKNKSITELNKKVRDMYNEVIEAQAKNNEIARAA 1295  
Qy 934 EQLLPKPGTHSEKSSSESANSK 956  
Db 1296 BELTKSQTHLDIVSEKDEKLAK 1318

RESULT 92  
S31132  
hypochemical protein F59B2.12 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Feb-1995  
C/Accession: S31132  
R/Sulston, J.; Du, Z.; Thomas, K.; Wilson, R.; Hillier, L.; Staden, R.; Halloran, N.; G  
awkins, T.; Ainscough, R.; Waterston, R.  
submitted to the ENBL Data Library, November 1991  
A/Description: The C. elegans sequencing project: A beginning.  
A/Accession number: S31132  
A/Status: preliminary  
A/Molecule type: DNA

A/Residues: 1-943 <SUL>  
A/Cross-references: EMBL:Z11505  
C/Genetics:  
A/Introns: 47/1, 132/3, 203/2, 433/3, 544/3, 586/3, 767/3

Query Match 3.3%; Score 170; DB 2; Length 943;  
Best Local Similarity 17.8%; Pred. No. 2.5;  
Matches 172; Conservative 149; Mismatches 360; Indels 284; Gaps 39;

Qy 25 PSLVKEPILKQTOASSISGADYAESGSKSLKLNKINETSQGVDDTVTDLFSDKRTTPEKIK 84  
Db 55 FAMPKLDASKAAMVHSSSHKHGHQSGSSSNTHSLTVVGADG----- 97  
Qy 85 DNLAGPREQELKAVTENTESEKQITSGSQLEQSKESL---SLNKTYTPSTSNWEICDFI- 140  
Db 98 -----KNITENSEKKOGYNKESKVDDEANETKIKSADGSVAASVQPMFCVPLC 145  
Qy 141 -----TKGNTLVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVASPAF 183  
Db 146 FLKLYTFHPSQIETGKSHNKSDDASSYGLEKSSKT-----Y 183  
Qy 184 TPDKKTAAIAYTSRAGENGESIQLDVDGKEITINEGEVFNYSYLLKKVTIPTGYKH- 237  
Db 184 ADKNGTMLSNTNKNQNSRGAALD-EGNEFVNQONADGTFLRNN---TGKNTDEHLS 238  
Qy 238 -----IGQDAFVDKNIAEAVNLPSLETISDYAFALALKQIDLDPNLK----- 281  
Db 239 HNVLDENQAQMSIGADGTSHTITNRKSGVGDSSHNAASD---AHSNPFESLDAQGNKKSQNY 295  
Qy 282 -----AIGELAFFNQI-----TGKLSLPROLMRLAERAFKSNHIKT----- 318  
Db 296 KKAASASGSVADFPESNLESLEKNADGTSMSNSTGNFNNTSYDKATAEEVMSKKNVADGTS 355  
Qy 319 ---TEFRGNSLKVIGEASFQDNDSQLMLPDGLEKIESEAFTEGPNPDGHDYNNRV-VLWTK 374  
Db 356 SMEASHAGNSKINSASGQSSDLS-MVGPNGIKSHSTSNKTDYALDEANQSAGSISEQ 414  
Qy 375 SGKPSGLATENTYVNDPKSLWQSPEDITYKWLBEPTYQKNSVTFGSKGLQKVRNK 434  
Db 415 IGRKQBSLNESSIESGRKAESRNTAADTLDSVDANGTVSSSHSKSASGTSLDE---NH 471  
Qy 435 NLEIPKQHN-GVTITEIDNAPRVDD--FONKTLRKYDLEEVKLPSTIRKIGAFAPASN- 490  
Db 472 N-----KTHAQASVDEHGNKKNHSIDGSYRNKKTGFEQNSE--MSAISKADGTMSQVNV 525  
Qy 491 ---NLKSFEASDDL-----BEIKEGAP-----MNNRI----- 514  
Db 526 KNDTNRTYEAKEKSALEKHNKESDGTGFKDES KGSNSRNVNRTDGSNLA VGSYVSGKGV 585  
Qy 515 ---ETLEKDKLVITIGDAAFHINHIY-----AIVLPESVQIEGRGAFRQNGANNLI-- 562  
Db 586 SSNETIASNAFNSTDAESNQFDHLHQKTANGTBITHAKDSKQVAASAKSSLDTSMSA 645  
Qy 563 --FMGSKV-KTLGMA-----PLSN-----RLEHLDLSEQ 589  
Db 646 VDAKGNKVDKTSQQAADSHDAISASSVDVDAKIVGHADRSSESINDSSNQTHASEHNDSSKQ 705  
Qy 590 KQ-----LTEIPVQAFSDNALKEVLLPASLKTIRBEAFKK-----NHLKQLEVAS 634  
Db 706 SEHEKRNQADGSPSDVSSNAKKNVAKDA-----TDVRQNAFANVDAAGNSVEVKNSL 759  
Qy 635 ALSHIAFNALDDNDCGEQFQKNVVKVTHNSYALADGEHFI VDPDKLS-STIVLEKILK 693  
Db 760 VENH-----SNDSDASSDSKIHKSVDGTF-ELTDAGHSNVSTSHLTGEOQAQIKKLA 811  
Qy 694 LIEG---LDYSTLRQTQTQFRDMTTAGKALLSK-----SNLRQGEKQKFLQEAQFFLGR 745  
Db 812 MADGSVVMGNDNSHITHEKSRADVDAAGHKAHLSKADSGSSTDIIDEG-----FKHH 860  
Qy 746 VDLDKAIATAKFAKLVTKKATKNGQLLERSINKAVLAYNNSAIKKNVVRLEKELDLITGL 805  
Db 861 SDLESRGEGAKQ---QRYOKLGNGETESSMD---VGYEKSMAGKGGDQTSHHK---TLA 909

Qy	676	VDPKLGSTIV---DLEKILKILIEGLD---YSTLRQTTQTQPRDMMTAGKALLSKSNLRQ	729
Db	615	EEAEAAAKEALKNDVNNAYEIVQGDGRYYVVLK-----IEVADEEPEP	658
Qy	730	GEKQKFFQIEAQFFIGRVDLDKAIKAEKALVTTKATKNGQLLERSINKAVLAYNNSAIKK	789
Db	659	GEDTPEVQS-----GYATYEEAEAAAKEA-----LKEDKVNNAVYEVVQGDGRY	702
Qy	790	ANVXRLEKELDLLGLVGEKGPLAQATWVQGV-----YLLKTP-----LPLPEYIIGLVNY	840
Db	703	YVVLKIEDKEDEQGEPEGNP-----GITDEWLLKNAKEDAIKELKEAGISSDIY	754
Qy	841	FD---KSGKLIYALD-MSPTIGEGOKDAYG-NPILNVDE-----	874
Db	755	FDAINKAKTVGEVGEALKNELKHAHEKPGENPGITIDEWLLKNAKEAAIKELKEAGITAE	814
Qy	875	-----DNEGYPHALAVATL---ADYEG-----LDIKTILMSKLSQLTSIRO--VP	913
Db	815	YLFNLINKAKTVGEVSESLKILKAHAEKPGENPGITIDEWLLKNAKEDAIKELKEAGIT	874
Qy	914	TAAVHRA-----GIFQATQNAAAEAEQLLPKPG-----THSEKSSSESSEANSKD	957
Db	875	SDIYFDAINKAKTTIEGV-EALKNELKHAHKDEEFGKPGEDKPKDGGDKPKEDKK	933
Qy	958	RGLOSNP-----KTNRG-----RHSAILPRTSGKGSFVYGILGYTSVALLSLITA	100
Db	934	PGEDKPKEDKKPGKTKDQSPNKKKAKLPKAGSEAIL-----TLAAALSTLAGAYV	986
Qy	1003	-IKKKK 1007	
Db	987	SLKKRK 992	
RESULT 94			
Protein F21J9.12 [imported] - Arabidopsis thaliana			
C:Species: Arabidopsis thaliana (mouse-ear cress)			
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004			
C:Accession: F86378			
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;			
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.;			
Nature 408, 816-820, 2000			
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.			
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.;			
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.			
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.			
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.			
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.			
A:Reference number: A86141; MUID:21016719; PMID:11130712			
A:Accession: F86378			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-1864 <STO>			
A:Cross-references: UNIPROT:Q9FYL7; GB:AE005172; NID:g9743334; PIDN:AAF97958.1			
C:Genetics:			
A:Gene: F21J9.12			
A:Map position: 1			
Query Match 3.3%; Score 170; DB 2; Length 1864;			
Best Local Similarity 19.2%; Pred. No. 7;			
Matches 222; Conservative 176; Mismatches 401; Indels 356; Gaps			
Qy	23	EVSLVKEPILKQTQASSISGADYAESGKS-----KUKINETSGPDDTVTDLFS	74
Db	333	EFEKLKAELELEKTKCTNTEKLSNAVTKGALVQNRDALKHQLSEKTTTELANRLTEL-Q	391
Qy	75	DKRTTPPEIKDNLAKGPREQLKAVTENT-ESEKQITSGSQLEQSKESLSLNK-----T	127
Db	392	EKETALR--SSEVMKGQLEQSL-----TEKTBDEKCYAEINDRKSVSLYELTKLEBOSL	446
Qy	128	VPSTSNWEICDFITKGNLT-LVLSKSGVGEK--LSQTDHLVLPQSQADGTLTIQVASFAT	184



```

Db      447 AEKTKELSEC--LTKQEMSTALDSGLDQKELASDAMVASYQ-----EMLSVN-SII 498
Qy      185 PDKTAIAE-YTSRAGENGESIQ-----LDVPGKEIINEGEVFNSEYLLKKVTIPTGYKHIG 239
Db      499 ENIETILSNIVTPBEGHFDIVKVRSLAERKELTNVQEVNR--LKDL-----546
Qy      240 QDAFVNDKNIABVNPESLEIS--DYAFALALQKIDLPDLNKAIGELAFDNIQTG-KL 297
Db      547 -----IVSIDLPEEMSSSLESRLAWRESFLOGKDEVNAL-----QNRIESVSM 591
Qy      298 SLPRQMLRAERAFKSNHKTIEPRGNSLKVIGEASFDNDLSQLMDPLGELKTEREAPT 357
Db      592 SLSEAE-----EKSNIKELDDLSF-----SLKMEETAEAR 623
Qy      358 QNPQDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQBSPEIDYTKWLEEDFTYQKN 417
Db      624 GSLEREIEVRLV-----ETSGLWTEGV-----EDHT--SS 652
Qy      418 SVTGESNKGLOKVKRNKLEIPKHNGVTIIEIGNAPRVND-----FQNTKIRKYDLEEV 473
Db      653 DINLLVDRSPDKIEKQ-----IRDSSDSSYGNEEIEFAFOS--LLYVRDLFFS 698
Qy      474 KLPSTIRKIGAFAPSNLKGFEASDDLE-EIKEGAFMNNRIETLE-----LK 520
Db      699 LCKEMLGEGELISFOVSNL-----SDELKIASQELAFVKEEKIALEKDLERSEKSAALLR 753
Qy      521 DKL---VTIGDAAFHINHIIYAVIPESVQIEGRAPR-----QNGANNIIFMGSK--- 567
Db      754 DKLSMAIKKGKGLVDREKFKTQLDKKEKSEIKLMLEQLQGLGTVGVKYNQIDMLSRDLE 813
Qy      568 -----VKTGEMAFSLNRHLHDLSEK-----OLTRIPVQAFSDNALKEV----- 608
Db      814 RTKSELETAVKBERDQLOQSLSLIDTLQKWKSVSEIIALPVDLASEDSEKIDRLAG 873
Qy      609 -LLPASLKTIRE-EAFKK-----NHLKQELVASALSHAFNALDDNDGDSQFQNKV 657
Db      874 YIQEVQIARVQEIEKVEKSEVDALTSLKLAETQTLAKLVEDALSTADNLSRLTEENRN 933
Qy      658 VVKTHNS-----YALADGEHFIYDPDKLSSTIVDLEKIL-----692
Db      934 VQAKENAELELOKAVADASSVASSELDEVLATKSTLEAALMQAERNISDIIEKEEAQGR 993
Qy      693 -----KLIEGLDYSTLRQTQT-----QFRD---MTT 716
Db      994 TATAEMEQLKQKASIQKNKLTEA--HSTINSLEETLAQTESNMDSLSKQIEDDKVLTT 1051
Qy      717 AGKALLSKNLRQGBKQKFLQEAQFFLGRVLDLQKAIKAEKAL-----759
Db      1052 SLKNELEKLTAEAFERNKMAEASLTI--VSHEEALMKAEKSLSALQEMVKABGEISTL 1109
Qy      760 -----VTKATKNGQLLERSI-----NKAVLAYNNSAIKKAN-----VKRLEKBL 799
Db      1110 SSKLVNCHCELLAGSGNSQSLSLEIITHLDNLQMLLKDGLLISKVNEFLQKPKSL-RDV 1168
Qy      800 DLL-----TGLVEKGFPLAQATWQGVYLLKTPPLPEYYIGLVNVPDKSGKLIY 849
Db      1169 DVIAIDITRNTGENGLLAGEMGNAEAVLL--ITLL-----YFQDDSTEAK 1211
Qy      850 AL--DMSPTTIGEGQKDAYGNPILNVDEN-----EGYHALAVATLADYEGL--DIK 896
Db      1212 SLLSDLSNVTEPENSQGS-----ADEDEISSLRKMAEGVRLNRKNTLENNPFGSTSID 1268
Qy      897 TLNLSKLSQLSIR--QVPTAAVHRAGI---FOAIONAAAEALQPLPKGTHSEKSSSES 952
Db      1269 TLIATLMQNTAARADVLNIVGHNSLSLEQVRSVENIVREQENTI--SALQKDLSSLISA 1326
Qy      953 ANSKDRGLQSNPKTN 967
Db      1327 CGAAARELOLEVKNN 1341

```

RESULT 95

```

S46773
myosin heavy chain - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YHR023w; type II myosin
C:Species: Saccharomyces cerevisiae
C>Date: 07-Oct-1994 #sequence revision 11-Aug-1995 #text_change 09-Jul-2004
C:Accession: S46773; S05806; S12323
R:Du, Z.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid 8082.
A:Reference number: S46773
A:Accession: S46773
A:Molecule type: DNA
A:Residues: 1-1928 <DUZ>
A:Cross-references: UNIPROT:P08964; EMBL:U10399; NID:G500689; PIDN:AAB68872.1; PID:G50006
R:Watts, F.Z.; Shiels, G.; Orr, E.
EMBO J. 6, 3499-3505, 1987
A:Title: The yeast MYO1 gene encoding a myosin-like protein required for cell division.
A:Reference number: S05806; MUID:88111539; PMID:3322809
A:Accession: S05806
A:Molecule type: DNA
A:Residues: 1-45, 'T', 47-58, 'S', 60-329, 331-342, 'S', 344-528, 'ARGHDR', 536-540, 'V', 542-598, 61
A:Cross-references: EMBL:X06187; NID:G4015; PIDN:CAA29550.1; PID:G4016
R:Sweeney, F.P.; Watts, F.Z.; Pocklington, M.J.; Orr, E.
Nucleic Acids Res. 18, 7147, 1990
A:Title: The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide sequence.
A:Reference number: S12323; MUID:91088308; PMID:2263482
A:Accession: S12323
A:Status: translation not shown; significant sequence differences
A:Molecule type: DNA
A:Cross-references: EMBL:X53947
C:Genetics:
A:Gene: SGD:MYO1
A:Cross-references: SGD:S0001065; MIPS:YHR023w
A:Map position: 8R
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; nucleotide binding; P-loop
F:78-779/Domain: myosin motor domain homology <MMOT>
F:180-187/Region: nucleotide-binding motif A (P-loop)
F:459-528/Domain: actin binding #status predicted <ACB>
F:186/Binding site: ATP (Lys) #status predicted

Query Match      3.3%; Score 170; DB 2; Length 1928;
Best Local Similarity 20.0%; Pred. No. 7.3;
Matches 200; Conservative 150; Mismatches 349; Indels 302; Gaps 47;

Qy      57 KINETSQGVDDTVTDLFSDKRTTPEKIKDNLAKPREQELKAVTETSESEKQITSGSOLE 116
Db      865 KFNEQINKKNDLQEMESKKKFLSEK-----NQKTVNELENTQ-----DLIN 906
Qy      117 QSKESLSLN-----KTVFSTSNWEICDFITKNTLVGLSKSGVEKLSQTDHLVLPQSA 169
Db      907 QEKENLRKNSLNRVKTSTSETLQKQFDLLVSEKDEI-----SREKLEVAQNLEEAHQK 960
Qy      170 ADGTQLIQVASPAFTPPDKTAIAEYTSRAGE-NGEISQLDQD-----GKEIINEGEVFN 222
Db      961 IQGLQ-----ETIRREATLEKSLKNNELIKQISDLNCDISKESQSSQSLIKESKLK 1013
Qy      223 SYLLKKVVTIPGYKHIGQDAFVDNKNIAEYNLPESLET-----ISDYAFALHAKQIDL 276
Db      1014 ENETKRLKDVINSKEEIKSFNDKLSSEEDLDIKVTLKKNKCNIAMSRLQSLVTENS DL 1073
Qy      277 ---PDNLKATGELAFPDNQITGKLSLPQLMRLAER-----309
Db      1074 RSKNENPKK--EKAALNNQLNKES---ELLKMKEDIHNNKELATTFSKQRDDAVSBHGK 1128
Qy      310 ---AFKSNHKTIEFRGNSLKVIGE-ASFO-----DNDLSQLMLPDGLKIESEA 355
Db      1129 ITAELKETRQLTIEYKSNYQKIEEYNFQRETKEQKKNLSLVESINDSKIELEARL 1188
Qy      356 FTGNPGDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQBSPEID-----403
Db      1189 SQEISLQNYLNMKRI-----SGNSVETNISSTRSTSYSDPDLDKEDIKKYYVDLQ 1238

```



A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-1881 <KUR>	
A:Cross-references: UNIPROT:Q9L7Q2; GB:AB005672; PIDN:AAK74809.1; PID:g14972138; GSPDB:G	
A:Experimental source: strain TIGR4	
C:Genetics:	
A:Gene: SP0664	
Query Match	
Best Local Similarity 3.3%; Score 169.5; DB 2; Length 1881;	
Matches 220; Conservative 176; Mismatches 443; Indels 309; Gaps 53;	
Qy	29 KEPILK-----QTQASSISGADYAESGKS-----KUKINETSPV 65
Db	388 KEPVDKSELNQIDKASSVSPDYASTASYNALGPVLETAAGVYASEPVKQPEVNSETNKL 447
Qy	66 DDTVDFSDKRTTPEKIKDNLAKGPRE-----QELKAVTENTSEKQITSGSQLEQSK 119
Db	448 KTAIDALNVDTKELNNTIADAKTKYEHYSDBSQNLQ--TEVTKAEK-VAANTDAQSE 504
Qy	120 ELSLUNKTVPTSNWEICDFTKGNLTVLGSKSGVEKLSQTDHLVLPQQAADGTOLIOVA 179
Db	505 VNEAVEKLTATIEK-----LVLESEKPILTLTSTDKKILERE-----VA 544
Qy	180 SPAFTPKKTAIAEYTSRAGENGHISQLDVGKEIINEGEVFNSEVLLKKVPTPTGYKHIG 239
Db	545 KYTLENQNKTKIKSITA-----ELKK-----GESEVINT-VLTTDDKVTETISAAPKNLE 593
Qy	240 --QDAFVQNKIAEYVNPLESLETISDYAFALHAKQIDLPDLNKLKAIGELAFDFDQITGKL 297
Db	594 YKYEYLTSTMYIDRGNGEETLENQNI-QLDLKKVEL-KNIKRT-DLIKYEN---QKE 647
Qy	298 SLPLQMLRAERAFKSNHKITIEFRGNSLKVIGEAFQDNLSQMLPDGLEKIESEAPT 357
Db	648 TNESLITTPDD--KSNYY-----LKIT-----SNNQKTTLL--AVKNIEBETTWN 688
Qy	358 GNPQDDHNNRVLTGSKGNPSGLATENTYVNPDKSLWQSPEDYTKWLEEDFTYQKN 417
Db	689 GTP-----YKVTADNLVSTADNKF-----ESEYHYIEKPKVHEDN 728
Qy	418 SVTGF-----SNKGLQKVKRNKNLEIPKQNG-----VTITEI 450
Db	729 VYNNFKELVEALQNDPSKEYRLQGSMSARNVVPNGKSYITKEFTKLSSEGKQPAITEL 788
Qy	451 GDNAF-----RNVDPONTKRLKYDLEEV-KLPSTIR-----KIGAPAFOSNNLK 493
Db	789 EHPFNVTNATINNVPNFENVEIERSQDNIASLANTMKGSSVITNVKITGTLSRNNVA 848
Qy	494 SPEASDDLEEIEKEGAFMNNRIETLELKDQKLVITIGDAAPHINHIYAIVLPEVQIEGRSAF 553
Db	849 GF-----VNNMNDG-----TRIENVAFFGKLHSTSGNGSHGTGIAG-----TNYRGIVRKAY 895
Qy	554 RQNGANNLIFMGSKVKT-----LCEMAFLSNRL--EHLDSLSEKQKLTETI 595
Db	896 -----VDATIGNKTRASLLVPKVDYGLTDLHLICTKALLTESVYVKGIDVSNPVEVCAI 950
Qy	596 -----PVOAFSDNALKEVLLPASLKTIR-EEAFKQNHKLQLEVASALSHIAFNALDDNDG 649
Db	951 ASKTWPVGTVGNSV-----SYAKIIRGEELFGSDVDSDSYASAHIKDLYAVEGYSSG 1003
Qy	650 DEQPDNVVVKTHNSYALADGEHEFVDPDKLSSITVLEKILKLI EGLDYSTLRQTTQT 709
Db	1004 NRSFRKSKTFTKLTKEQADAKVTTFNTITADKLES---DLSPFLAKNEEKAYSSI-QDYNA 1059
Qy	710 QPRDMTTAGKALLSKSN-----LRQCEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKAT 765
Db	1060 EYNQAYKNLEKLI PPNKYDVIYQGNK-----LNKEHHLNTEKVLVSVM 1104
Qy	766 KNGQLLER--SINKAVLAY-----NNSAIKKANVXR-----LEKEL 799
Db	1105 NNNEFITNLDANKIIVHYADGTQDYFNLSSESSEGLSNVKEVTTITDLGIKYTPNIVQKD- 1163
Qy	800 DLLTGLVSGKGLAQTAVQGVYLLKTLPLPEYYIGL--NVYFPKS-----GKLIY 849
Db	1164 --NTTLVNDIKSILESVELSQTYQHNLNLDGTRVNAIKDLYLEESFTDVKENLTNLIT 1221
Qy	850 ALDMSDTIGEGQKDAYGNPILNVDEDEGEGHALAVATLADYEG-----LDIKTIL----- 899
Db	1222 KLVQNEEHQLNDSPAARQMIRDKVEKNKAALLGLTYLNRYYGKVGFDVNIKELMLFKPD 1281
Qy	900 --NSKLSQTSIROVPT-----AAHHRAGIFQAIONAAEA----- 933
Db	1282 FYGKSVSLDRLEBIGSKENNIKGRSRTFDAGQVLAKYTKSGNLDLAFNLNRLQFTNIDN 1341
Qy	934 --EQLLPKPGTH---SEKSSS-SESANSKORGLSQNPKTNRGRHSAILPRGTSGKSGFVYG 987
Db	1342 MNDWFIDATEDHVIAERASEVEEIKNSKHRAF-DNLKRSRLRNT-ILPLLNIIDKAHLYL 1399
Qy	988 ILGYTSA 995
Db	1400 ISNTNAIA 1407
RESULT 98	
T01362	
probable myosin heavy chain At2g34730 - Arabidopsis thaliana	
N:Alternate names: hypothetical protein T29F13.6	
C:Species: Arabidopsis thaliana (mouse-ear cress)	
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004	
C:Accession: T01362; C84760	
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul	
submitted to the EMBL Data Library, May 1998	
A:Description: Arabidopsis thaliana chromosome II BAC T29F13 genomic sequence.	
A:Reference number: 21479	
A:Accession: T01362	
A:Status: translated from GB/EMBL/DBJ	
A:Molecule type: DNA	
A:Residues: 1-829 <ROU>	
A:Cross-references: UNIPROT:O64584; EMBL:AC003096; NID:g3132469; PID:g3132472	
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;	
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.	
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.	
Nature 402, 761-768, 1999	
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.	
A:Reference number: A84420; MUID:20083487; PMID:10617197	
A:Accession: C84760	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-829 <STO>	
A:Cross-references: GB:AE002093; NID:g3132472; PIDN:AAC16261.1; GSPDB:GN00139	
C:Genetics:	
A:Gene: At2g34730; T29F13.6	
A:Map position: 2	
A:Introns: 728/2; 770/3	
Query Match	
Best Local Similarity 3.3%; Score 169; DB 2; Length 829;	
Matches 177; Conservative 155; Mismatches 328; Indels 304; Gaps 45;	
Qy	28 VKPEILKQTOASSISGADYAESGSKLKN-----ETSGFVDDTVTDL-----FSDKR 77
Db	1 MEEREVKSNGS-----LEFHDDTLSSSLQVGVLENENPDVDFLEDLSYWEINDRL 55
Qy	78 TTPBKIKDNLAKGPREELKAVTENTSEKQITSGSQLEQSKESLSL-----NKTVPSTSN 133
Db	56 TISRVSDSIIRG-----MVTAIESDAAEKIAQDXLELSKIRETLLLYHYVGEENSES 110
Qy	134 WEICDFTKGNLTVLGSKSGVEKLSQTDHLVLPQQAADGTOLIOVASFAFTPKKTAIAE 193
Db	111 RLHDEUTQG-----SSSLKKKARQKQLMLVEELTNREYTHINGSATVD----- 157
Qy	194 YTSRAGENGEISQLDVGKEIINEGEVFNSEY--LKKVTIPTGYKHIGQDAFVNQKIAE 251
Db	158 -----DSLGLDSSPHETRSKTVDKMDLSKSIL-----ETVLKRN--D 194

QY 252 VNLPSLETISDYAPAHALALQOIDLDPDLNKAIGELAPFDNOITGKLSLPLQMLRAERAF 311  
DB 195 MELPSSWQOEHDFO-----KEIE-----SAV-----VTSVL-----RSL 223  
QY 312 KSNHIKIT-----BFRGSLKVIGEASQPDNDLSQLM-LPDGLEKIESEAFGNPGDD-- 363  
DB 224 KDEYFORLLDQKAEGGNRSILIG-----NIKEITGLRQELFAIRKSFLDHENGDEAG 276  
QY 364 -----HYNNRVVLWT--KSGKNPSGLATENT-----YVNPDKSLWQESP 400  
DB 277 EVGDRKRVQLHRKMSGSLNSVSSWENGKEESSTGLIPEHNETLRHMSPDEMI----- 331  
QY 401 EIDYTKWLEEDFTYKQSVTFGSNGKLGQVKRNKNLEIPK-----QHNGVITITEIGD 452  
DB 332 -----NHFKIEMN-----KMKRDHDYKIQELTEQCFTFKRYLNLTERGS 371  
QY 453 NAFNRVDPQNKTLRKYLDEEVKLPSTIRKIGAFAFQSNLNKSFASDD--LEELKEGAFM 510  
DB 372 FSVFGKDKELGALKK-----KIPPVISKDLKILMEDEKFPVS-EKNDAGLKRQLDSILL 424  
QY 511 NNRITLLEKDKLVITIGDAAFHINH-----IYAI 539  
DB 425 ENR-----QLKDSL-----SDAAEKMSQLSQAEADHOELIRKLETVDVEDSRNEASIYEDVYGC 477  
QY 540 VLPESVQEIGRSAFRQNGANNLI FMGSKVYTLGEMAFLSNR-----LEHLDLSEQKQLTEI 595  
DB 478 FVTBFVGQI-----KCTQOETDLEHSLMLREAYELLLEDLARKEARKSKB- 521  
QY 596 PVQAFSONALKEVLLPASLTKIRBEAFKKNHLK-----OLEVASALSHIAFNALDNDGDE 651  
DB 522 --DFEDSCVKSVMMECCSVIYKEAVKEAHKIVELNLHVTEKEGTLRSEVVDKERLKE 578  
QY 652 QF-----DNKVYVVKTHNSYALADGEHFIVD-----PDKLSSTIV 686  
DB 579 EIHRLGCLVKEKENLVQTAENNLATERKKIEBWSQOINDLQSQVERQETIQDKIEALS 638  
QY 687 DLKILKILIEGLD--YSTLR--QTTOQFDMTTAGKALLSKNLKRGQEKQKFLQEAQF 741  
DB 639 VSARELEKVKYETKISLSRELELARELSUKEMDEKRTKEETKAEKETUKKQL-- 696  
QY 742 FLGRVDDLKAIYAK-----AFKALVTKKATKNGQL-----LERSINKA---VLA 781  
DB 697 ---VSLDLVVPPIIKGFIDILEGLIAEKTQKTNRLKNMQSQLSDLSHQINEVKGRAST 752  
QY 782 YNNSAIKK-ANVKLEKELDLTLGLVEGKGPLAAQWVGVL-LKTPLPUPPEYIG--- 836  
DB 753 YKORLEKKCCDLKKAEEVDLLGDEVE-----TLDLLEKITYIALDHYSPILKHPYGVST 807  
QY 837 LNVY 840  
DB 808 ISVY 811

RESULT 99  
T13174  
gpi50 protein - fruit fly (Drosophila melanogaster)  
C/Species: Drosophila melanogaster  
C/Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
C/Accession: T13174  
R/Tian, S.S.; Zinn, K.  
J. Biol. Chem. 269, 28478-28486, 1994  
A:Title: An adhesion molecule-like protein that interacts with and is a substrate for a  
A:Reference number: Z17630; MUID:95050638; PMID:7961789  
A:Accession: T13174  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1051 <TIA>  
A:Cross-references: UNIPROT-Q24007; EMBL:U15220; NID:G595859; PID:G595860; PIDN:AAA61798  
A:Experimental source: strain Oregon R  
C/Genetics:  
A:Cross-references: FlyBase:FBgn0013272  
A:Map position: 2

QY 252 VNLPSLETISDYAPAHALAKQIDLPDNLKAIGELAFPDNQITGKLSLPQMLRAERAF 311  
DB 195 MELPSSWQOEHDFO-----KEIE-----SAV-----VTSVL-----RSL 223  
QY 312 KSNHIKIT-----BFRGSLKVI GEASQPDNDLSQML-LPDGLEKIESEAF TGNPGDD-- 363  
DB 224 KDEYFORLLDQKAEFGNRSILIG-----NIKEITGLRQELFAIRKSFLDHENGDEAG 276  
QY 364 -----HYNRRVLWT--KSGKNPSGLATENT-----YVNPDKSLWQESP 400  
DB 277 EVGDRKRVQLHRKMSGSLNSVSSWENGKEESSTGLIPEHNETLRHMSPDEMI----- 331  
QY 401 EIDYTKWLEEDFTYKQSVTFGSNGKLGQVKRNKNLEIPK-----OHNGVITITEIGD 452  
DB 332 -----NHFKIEMN-----KMKRDHDYKIQELTEQCFTFKRYLNLTERGS 371  
QY 453 NAFNRVDPQNKTLRKYLDEEVKLPSTIRKIGAFAFQSNLNKSFASDD--LEELKEGAFM 510  
DB 372 FSVFGKDKELGALKK-----KIPPVISKDLKILMEDEKFPVS-EKNDAGLKRQLDSILL 424  
QY 511 NNRITLLEKDKLVITIGDAAFHINH-----YIAY 539  
DB 425 ENR-----QLKDSL-----SDAAEKMSQSAEADHOELIRKLETVDVEDSRNEASIYEDVYGC 477  
QY 540 VLPESVQEIGRSAFRQNGANNLI FMGSVKVTLGEMAFLSNR-----LEHLDLSEQKQLTEI 595  
DB 478 FVTEFVGQI-----KCTQOETDLEHSLMLREAYELLLEDLARKEARKSKB- 521  
QY 596 PVQAFSONALKEVLLPASLTKIRBEAFKKNHLK-----OLEVASALSHIAFNALDNDGDE 651  
DB 522 --DFEDSCVKSVMMECCSVIYKEAVKEAHKIVELNLHVTEKEGTLRSEVMDKERLKE 578  
QY 652 QF-----DNKVVVKTHNSYALADGEHFIVD-----PDKLSSTIV 686  
DB 579 EIHRIGCLVKEKENLVQTAENNLATERKKIEBWSQOINDLQSQVERQETIQDKIEALS 638  
QY 687 DLKILKILIEGLD--YSTLR--QTTOQFDMTTAGKALLSKSNLRQGEKQKFLQEAQF 741  
DB 639 VSARELEKVKYETKISLSRELELARELSUKEMDEKRTKEEKUSETKAKEETUKKQL-- 696  
QY 742 FLGRVDDLDAKAYK-----AFKALVTYKATKNGQL-----LERSINKA---VLA 781  
DB 697 ---VSLDLVVPPIIKGFIDLEGLIAEKTQKTNRLKNMQSQLSDLSHQINEVKGRAST 752  
QY 782 YNNSAIKK-ANVKLEKELDLTLGLVEGKGPLAQTWVGVT-LKTPLPUPPEYIG--- 836  
DB 753 YKORLEKKCCDLKKAEEVDLLGDEVE-----TLDLLEKITYIALDHYSPILKHPYGVST 807  
QY 837 LNVY 840  
DB 808 ISVY 811

RESULT 99  
T13174  
gp150 protein - fruit fly (Drosophila melanogaster)  
C/Species: Drosophila melanogaster  
C/Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
C/Accession: T13174  
R/Tian, S.S.; Zinn, K.  
J. Biol. Chem. 269, 28478-28486, 1994  
A:Title: An adhesion molecule-like protein that interacts with and is a substrate for a  
A:Reference number: Z17630; MUID:95050638; PMID:7961789  
A:Accession: T13174  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1051 <TIA>  
A:Cross-references: UNIPROT-Q24007; EMBL-U15220; NID:G595859; PID:G595860; PIDN:AAA61799  
C/Genetics:  
A:Experimental source: strain Oregon R  
A:Cross-references: FlyBase:FBgn0013272  
A:Map position: 2

CjDate: 12-Oct-1989 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
CjAccession: A32491  
RjGeorge, E.L.; Ober, M.B.; Emerson Jr., C.P.  
Mol. Cell. Biol. 9, 2957-2974, 1989  
A>Title: Functional domains of the Drosophila melanogaster muscle myosin heavy-chain gene  
AjReference number: A32491; MUID:89384556; PMID:2506434  
AjAccession: A32491  
A>Status: Preliminary  
AjMolecule type: DNA; mRNA  
AjResidues: 1-2385 <GEO>  
AjCross-references: UNIPROT:P05661; GB:M61229; GB:M27194  
AjNote: the authors translated the codon TGC for residue 329 as Ser  
CjGenetics:  
AjGene: FlyBase:Mhc  
AjCross-references: FlyBase:FBgn0002741  
CjSuperfamily: myosin heavy chain; myosin motor domain homology  
CjKeywords: ATP; nucleotide binding; P-loop  
Fj137-1032/Domain: myosin motor domain homology #status atypical <MMO>  
Fj227-234/Region: nucleotide-binding motif A (P-loop)

Query Match 3.3%; Score 169; DB 2; Length 2385;  
Best Local Similarity 20.0%; Pred. No. 11;  
Matches 216; Conservative 154; Mismatches 418; Indels 292; Gaps 47;

QY 10 LTLTVSVVTHNOEVSFLVKBPILAKQTQASSISG--ADYAESSCKSKLKINETSGPYDD 67  
DB 1483 LTOEAVDLERNKKE---LEQTIQRKDKELSSITAKLEDEQVVVLKHQRQIKELQARIE 1539  
QY 68 TVTDLFSDK--RTTPEKIDNLAKPRQEKLKAVNTESSEKQITSGSQLEQSK---ES 121  
DB 1540 LEEVEAEERQAKAKQADLAR-----ELBELGERLEEAGGATS-AQIELNKKREAE 1593  
QY 122 LSLNTVPSTSNWEICDITKNTLVLSK-----SGVEKLSQTDHLVLPQQAADGTQLIQ 177  
DB 1594 SKLRDLLEAN-----IQHSTLANLRKKGNDVAEAMAEQVDQ----- 1632  
QY 178 VASFAFTPDKKTAAIEYTSRAGENSEISQ-----LQVDCKEIINEGEVENSYLLKVTIPT 233  
DB 1633 -----NKLKAKAEHRDQCHN-ELNQTTRTACDQLGRDKAEKEK--NEYYGQNDLRA 1681  
QY 234 GYKHTGQDAFVDNK-----NIAEVLNPESELETISDYAFALHAKQIDLPNLKAI 284  
DB 1682 GVDHITNEKAAQEKIAKQLOHTLNEVQSKLDNETNLDPDASKKL-SIENSLLRQLE 1740  
QY 285 ELAFTDNQITG-KLSLPRQL---MRLAERAFKSNHIKTIETFRGNSLKVIGASPDNDLS 340  
DB 1741 EAEQVSQLSKIKISLTQLEDTKRLADESRER-----ATLLGKFRNLEHLD 1789  
QY 341 QLMPLDGLKTESRAFTGNPGDDHYNRVVLWTKSGKNPSGLATENTYVNPDKSLMQ--- 397  
DB 1790 NLR-----EQVEEEA-----EGK--ADLQRLSKANAEQVWRKY 1823  
QY 398 ESPMIDYTKWLEE-----DFTYKNSVTGFSNK--GLQKVKR-----NKNLEIPKQ 441  
DB 1824 ESDGVARSEELAEAKKLQARLAEABETIESLNQKICIGLEKTKQRLSTEVEDLQLEVDR 1883  
QY 442 HNGVTITIGNAFNRV-----DFQNTKTRKYDLEEVKLPSTIRKIGAPA 486  
DB 1884 NAIANAEEKKQAFDKIIGEMKLVDDLAALDASQKCRNYSYTELFRLK-----GAYE 1937  
QY 487 FQSNLTKSFEASDLEELKEGAFPMNRITETLEKDKLVITIGDAAPHINHIYAVILPESVQ 546  
DB 1938 -----EQGEQLEAVRR-----ENKNLADEVKDLDDQIGEGG-----RNIIH 1972  
QY 547 EIGRSAPRQNGANNLIFMGSKVKTIGEMAFLSNRLHLDLSEK-----QLTEIPVQAFSD 602  
DB 1973 EIEKARKLEAKDEL-----QAALAEAEAELEQENKVLRAQLSELQSVRQID 2021  
QY 603 NALKEVLLPASLKTIRREAF---KQHLKQLEVASALSIAFNALDDNDGDEQPDNKVVV 659  
DB 2022 RRIQ-----KEBEFENTRKNQR-----ALDSMQASLEAEAKGA 2057  
QY 660 KTHNSYAL-ADGEHFIVDPDKLSSTIVDLEKILKLEGLDYSLTRQTQTQFRDMTTA- 717

Search completed: August 28, 2005, 10:49:18  
Job time : 99 secs

Db 2058 EALRMKKKLEADINELIALDHANKANAQAQNIK-----RYQQQLKDIQTAL 2105  
QY 718 -----GKALLSKNLRQGEKQKFLQBAQPFGRVLDLKDATAKAEKALVTTKATKNG 768  
DB 2106 BEEQRARDAREQLGISERRANALQNELEERSRTLEQADRRGRQAQEQE-----LADAHE 2159  
QY 769 QLLERSINKAVLAYNNSAIKK---ANVKRLEKELDILLTGLVEGKGPLAQATWQGVILL- 824  
DB 2160 QLNEVSAQNASI-----SAAKRKLESELQTLHSDLDLELNEAKNSEKAKKAMVDAARLAD 2215  
QY 825 -----KTPLPLPEYIIGLVYFD-----KSG-KLIIVALDMS-----DTIGE 859  
DB 2216 ELRAEQDHAQTOEKLKRLAEQOIKEQLVRLDEAEANALKGGKKAIOKLEQVRVRELELD 2275  
QY 860 QOKDAYGNPILNVVDNEGTHALAVATLADYEGLDIKTILNSKLSQ--LTSIRQVPTA-- 915  
DB 2276 GEQRRHADAQKNLRKSERRVKELSFQSEEDRKNERHQDLVDKLOQKIKTKYKRQIEEAE 2335  
QY 916 -AYHRAGIFQAIQNAABAEQOLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAI 974  
DB 2336 IAAIINTLAKFRKAQQQLEFEAE-----BRADLAEQAISKFRAGKRGASVGRGASPAI 2385

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:48:37 ; Search time 169 Seconds  
(without alignments)

2343.607 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 5080

Sequence: 1 MKHLKTVALLTIVSVVTH.....LGYTSVALLSLTAKKKY 1008

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1767149 seqs, 392926209 residues

Total number of hits satisfying chosen parameters: 1767149

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 110 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US11A\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5080	100.0	1008	14	US-10-078-531-2
2	5076	99.9	1008	16	US-10-474-792-46
3	4901	96.5	971	14	US-10-078-531-7
4	4876	96.0	971	14	US-10-078-531-6
5	4871	95.9	969	14	US-10-078-531-8
6	4867	95.8	970	14	US-10-078-531-4
7	4831	95.1	963	14	US-10-078-531-5
8	4775	94.0	951	14	US-10-078-531-3
9	3777.5	74.4	1055	14	US-10-091-007-24
10	3777.5	74.4	1055	16	US-10-476-614-2
11	241.5	4.8	669	15	US-10-369-493-11351
12					Sequence 11352, A
13					Sequence 2948, Ap
14					Sequence 254, App
15					Sequence 410, App
16					Sequence 4474, Ap
17					Sequence 7646, Ap
18					Sequence 62, Appl
19					Sequence 135452, A
20					Sequence 86, Appl
21					Sequence 70177, A
22					Sequence 5834, Ap
23					Sequence 81, Appl
24					Sequence 12996, A
25					Sequence 440, App
26					Sequence 70580, A
27					Sequence 23, Appl
28					Sequence 4098, Ap
29					Sequence 43762, A
30					Sequence 43811, A
31					Sequence 5703, Ap
32					Sequence 12611, A
33					Sequence 321, App
34					Sequence 2, Appli
35					Sequence 2, Appli
36					Sequence 2, Appli
37					Sequence 2, Appli
38					Sequence 2, Appli
39					Sequence 2, Appli
40					Sequence 2, Appli
41					Sequence 2, Appli
42					Sequence 2, Appli
43					Sequence 2, Appli
44					Sequence 2, Appli
45					Sequence 2, Appli
46					Sequence 2, Appli
47					Sequence 2, Appli
48					Sequence 2, Appli
49					Sequence 2, Appli
50					Sequence 2, Appli
51					Sequence 2, Appli
52					Sequence 2, Appli
53					Sequence 2, Appli
54					Sequence 2, Appli
55					Sequence 2, Appli
56					Sequence 2, Appli
57					Sequence 2, Appli
58					Sequence 2, Appli
59					Sequence 2, Appli
60					Sequence 2, Appli
61					Sequence 2, Appli
62					Sequence 2, Appli
63					Sequence 2, Appli
64					Sequence 2, Appli
65					Sequence 2, Appli
66					Sequence 2, Appli
67					Sequence 2, Appli
68					Sequence 2, Appli
69					Sequence 2, Appli
70					Sequence 2, Appli
71					Sequence 2, Appli
72					Sequence 2, Appli
73					Sequence 2, Appli
74					Sequence 2, Appli
75					Sequence 2, Appli
76					Sequence 2, Appli
77					Sequence 2, Appli
78					Sequence 2, Appli
79					Sequence 2, Appli
80					Sequence 2, Appli
81					Sequence 2, Appli
82					Sequence 2, Appli
83					Sequence 2, Appli
84					Sequence 2, Appli



```

85 185 3.6 3917 14 US-10-171-311-8 Sequence 8, Appli
86 185 3.6 3925 14 US-10-171-311-6 Sequence 6, Appli
87 185 3.6 5795 9 US-09-815-242-13610 Sequence 12610, A
88 184.5 3.6 1479 16 US-10-437-963-106122 Sequence 106122,
89 184 3.6 891 15 US-10-369-493-13477 Sequence 13477, A
90 184 3.6 1665 15 US-10-282-122A-71690 Sequence 71690, A
91 183 3.6 921 17 US-10-732-923-3305 Sequence 3305, Ap
92 183 3.6 1091 15 US-10-369-493-6328 Sequence 6328, Ap
93 183 3.6 1192 15 US-10-282-122A-57178 Sequence 57178, A
94 183 3.6 1440 16 US-10-437-963-168983 Sequence 168983,
95 182.5 3.6 944 9 US-09-815-242-11391 Sequence 11391, A
96 182.5 3.6 944 15 US-10-282-122A-58864 Sequence 58864, A
97 182 3.6 1192 9 US-09-815-242-10903 Sequence 10903, A
98 181.5 3.6 1786 9 US-09-742-096-3 Sequence 3, Appli
99 181.5 3.6 1787 15 US-10-415-253-2 Sequence 2, Appli
100 181.5 3.6 2026 9 US-09-801-368-86 Sequence 86, Appli
101 181.5 3.6 2026 15 US-10-369-493-22175 Sequence 22175, A
102 181.5 3.6 5171 16 US-10-408-765A-2687 Sequence 2687, Ap
103 180.5 3.6 1202 9 US-09-864-761-43061 Sequence 43061, A
104 180.5 3.6 1221 15 US-10-335-977-7813 Sequence 7813, Ap
105 180.5 3.6 1992 15 US-10-369-493-6527 Sequence 6527, Ap
106 178.5 3.5 995 9 US-09-864-761-49017 Sequence 49017, A
107 178 3.5 1212 9 US-09-815-242-5279 Sequence 5279, Ap
108 178 3.5 1224 18 US-10-617-320-4347 Sequence 4347, Ap
109 178 3.5 1482 9 US-09-815-242-12484 Sequence 12484, A
110 177.5 3.5 2076 9 US-09-815-242-5815 Sequence 5815, Ap
```

ALIGNMENTS

```

RESULT 1
US-10-078-531-2
; Sequence 2, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE OF INVENTION: DNA FRAGMENTS
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1008
; TYPE: PRP
; ORGANISM: Streptococcus pyogenes
US-10-078-531-2

Query Match 100.0%; Score 5080; DB 14; Length 1008;
Best Local Similarity 100.0%; Pred. No. 1e-303;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKCHLTKVALTTLTVSVVTHNQEVPSLVKEPILTKQTQASSISGADYAESGSKLKLINE 60
DB 1 MKCHLTKVALTTLTVSVVTHNQEVPSLVKEPILTKQTQASSISGADYAESGSKLKLINE 60

QY 61 TSGPVDVDTVTLFSDKRTTPKIKDNLAGPREQELKAVTENSEKQITSGSQLEQSK 120
DB 61 TSGPVDVDTVTLFSDKRTTPKIKDNLAGPREQELKAVTENSEKQITSGSQLEQSK 120

QY 121 SLSLNTVPSTNWEICDFITKGNLTVLGSKSGVEKLSQTHLVLPQAAOGTQLIQVAS 180
DB 121 SLSLNTVPSTNWEICDFITKGNLTVLGSKSGVEKLSQTHLVLPQAAOGTQLIQVAS 180

QY 181 FAFTPDKKTAETYSRAGENGESQLDVGKGEIINEGEVFNLSYLLKKVTIPTGYKHIGQ 240
```

```

181 FAFTPDKKTAETYSRAGENGESQLDVGKGEIINEGEVFNLSYLLKKVTIPTGYKHIGQ 240
241 DAFVNDKNIAEVNLPESLETISDYAFALHALKQIDLDPNKAIGELAFDNQITGKLSLP 300
241 DAFVNDKNIAEVNLPESLETISDYAFALHALKQIDLDPNKAIGELAFDNQITGKLSLP 300
301 RQLMLRAERAPKSNHIIKTIEPRGNSLVKIGRASQDNDLSQMLPDGLEKIESAFTGNP 360
301 RQLMLRAERAPKSNHIIKTIEPRGNSLVKIGRASQDNDLSQMLPDGLEKIESAFTGNP 360
361 GDDHYNRVVLWTKSGKNPSGLATENTYVNPDKSLMQESPEIDYTKMLEEDFTYQKNSVT 420
361 GDDHYNRVVLWTKSGKNPSGLATENTYVNPDKSLMQESPEIDYTKMLEEDFTYQKNSVT 420
421 GFSNKGKLVKVRNKNLEIPKQHGVTITEIGDNAFRNVDFQNKTLRKYDLEVKLPSTIR 480
421 GFSNKGKLVKVRNKNLEIPKQHGVTITEIGDNAFRNVDFQNKTLRKYDLEVKLPSTIR 480
481 KIGAFAPQSNNLKSFASDDLEETKEGAFMNNRIETLELKDQKLVITIGDAAPHINIIAIV 540
481 KIGAFAPQSNNLKSFASDDLEETKEGAFMNNRIETLELKDQKLVITIGDAAPHINIIAIV 540
541 LPESVQETIGRSAPFRQNGANNLI FMGSKVKVTIGEMAFLSNRLEHLDLSQKOLTPIVQAF 600
541 LPESVQETIGRSAPFRQNGANNLI FMGSKVKVTIGEMAFLSNRLEHLDLSQKOLTPIVQAF 600
601 SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSIAFNALDDDDGDFQFNKVVVK 660
601 SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSIAFNALDDDDGDFQFNKVVVK 660
661 THNSYALADGEHPIVDPDKLSSTIVLEKILKILIEGLDYSTLTQTTQTQPRDMTTAGKA 720
661 THNSYALADGEHPIVDPDKLSSTIVLEKILKILIEGLDYSTLTQTTQTQPRDMTTAGKA 720
721 LLSKSNLRQGEKQKPLQEAQFPLGRVDLDKAIKAERKALVTKKATKQGLLERSINKAVL 780
721 LLSKSNLRQGEKQKPLQEAQFPLGRVDLDKAIKAERKALVTKKATKQGLLERSINKAVL 780
781 AYNNSAIKCANVKKLEKELDLTLGLVEGKGPLAQATWVGQVYLLKTPPLPEYIIGLVNY 840
781 AYNNSAIKCANVKKLEKELDLTLGLVEGKGPLAQATWVGQVYLLKTPPLPEYIIGLVNY 840
841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDSDNNGYHALAVATLADYEGLDIKTILN 900
841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDSDNNGYHALAVATLADYEGLDIKTILN 900
901 SKLSQLTSIROVPTAAVHRAGIFQAIQNAAAEAQQLPKPGTHSEKSSSSSANSKORGL 960
901 SKLSQLTSIROVPTAAVHRAGIFQAIQNAAAEAQQLPKPGTHSEKSSSSSANSKORGL 960
961 QSNPKTNRGRHSAILPRTGSKGSFVYGILGYSVALLSLITAIKKKKY 1008
961 QSNPKTNRGRHSAILPRTGSKGSFVYGILGYSVALLSLITAIKKKKY 1008

RESULT 2
US-10-474-792-46
; Sequence 46, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Wintner, Lourie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474,792
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 46
; LENGTH: 1008
```

; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes  
; US-10-474-792-46.

Query Match 99.9%; Score 5076; DB 16; Length 1008;  
Best Local Similarity 99.9%; Pred. No. 1.8e-303;  
Matches 1007; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKGHLKVALTLTVSVVTHNQVFSLVKPEILKQTQASSISGADYABSSGSKLKINE 60  
DB 1 MKGHLKVALTLTVSVVTHNQVFSLVKPEILKQTQASSISGADYABSSGSKLKINE 60

QY 61 TSGPVDVDTVDLPSDKRTTPEKIKONLAKGPREQELKAVTENTSEKQITSGSQBSKE 120  
DB 61 TSGPVDVDTVDLPSDKRTTPEKIKONLAKGPREQELKAVTENTSEKQITSGSQBSKE 120

QY 121 SLSLNKTVSTSNWEICDFTTKGNTLVGLSKSVEKLSQTDHLVLPQADGTQLIQVAS 180  
DB 121 SLSLNKTVSTSNWEICDFTTKGNTLVGLSKSVEKLSQTDHLVLPQADGTQLIQVAS 180

QY 181 FAFTPDKTAIAEYTSRAGENGESIQDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240  
DB 181 FAFTPDKTAIAEYTSRAGENGESIQDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240

QY 241 DAFVDMKNIAEVLNPSLETISDYAPFAHLAKQIDLPDNLKAIGELAFDNOITGKLSLP 300  
DB 241 DAFVDMKNIAEVLNPSLETISDYAPFAHLAKQIDLPDNLKAIGELAFDNOITGKLSLP 300

QY 301 ROLMRLAERAFKSNHKTIEPRGNSLKVIAGEASFQNDLSQMLPDGLEKIESEAPTGNP 360  
DB 301 ROLMRLAERAFKSNHKTIEPRGNSLKVIAGEASFQNDLSQMLPDGLEKIESEAPTGNP 360

QY 361 GDDHYNRNVLLTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420  
DB 361 GDDHYNRNVLLTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

QY 421 GFSNKGLOKVRKNLLEIPKOHNGVTITEIGNAFRNVDFQNKTLRKYLDEEVKLPSTIR 480  
DB 421 GFSNKGLOKVRKNLLEIPKOHNGVTITEIGNAFRNVDFQNKTLRKYLDEEVKLPSTIR 480

QY 481 KIGAFAPQSNLKSPEASDDLEIEKEGAFMNNRIETLEKDKLVITGDAAFHNIHIAIV 540  
DB 481 KIGAFAPQSNLKSPEASDDLEIEKEGAFMNNRIETLEKDKLVITGDAAFHNIHIAIV 540

QY 541 LPESVQIEGRSAFRONGANLIFMGSKVKTGEMAFLSNRLHLDLSEQKQTEIPVQAF 600  
DB 541 LPESVQIEGRSAFRONGANLIFMGSKVKTGEMAFLSNRLHLDLSEQKQTEIPVQAF 600

QY 601 SDNALKEVLLPASLKTIREEAFKQHLKQLEVASALSHIAFNALDNDGDGEQFNKVVVK 660  
DB 601 SDNALKEVLLPASLKTIREEAFKQHLKQLEVASALSHIAFNALDNDGDGEQFNKVVVK 660

QY 661 THNSVALADGSHFIVDPDKLSSTIVDLKILKLEGLDYSTRQTTQFRDMTTAGKA 720  
DB 661 THNSVALADGSHFIVDPDKLSSTIVDLKILKLEGLDYSTRQTTQFRDMTTAGKA 720

QY 721 LLSKSNLRQGEKQKLOEAFPLGRVLDLKAIAKAEKALVTKATNGOLLERSINKAVL 780  
DB 721 LLSKSNLRQGEKQKLOEAFPLGRVLDLKAIAKAEKALVTKATNGOLLERSINKAVL 780

QY 781 AYNNSAIKANVRLEKELDLTLGLVEGKPLAQATWQGVYLLKTPLPPEYIYGLNYY 840  
DB 781 AYNNSAIKANVRLEKELDLTLGLVEGKPLAQATWQGVYLLKTPLPPEYIYGLNYY 840

QY 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDENEGYHALAVATLADYEGLDIKTIIN 900  
DB 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDENEGYHALAVATLADYEGLDIKTIIN 900

QY 901 SKLSQITSIRQVPTAAHYRAGIFQAIQNAABABQLLPKPGTHSEKSSSESANSKRG 960  
DB 901 SKLSQITSIRQVPTAAHYRAGIFQAIQNAABABQLLPKPGTHSEKSSSESANSKRG 960

QY 961 QSNPKTNRGRHSAILPRTGSKGSFVYIGLGYTSVALLSIITAIKKKKY 1008

DB 961 QSNPKTNRGRHSAILPRTGSKGSFVYIGLGYTSVALLSIITAIKKKKY 1008

RESULT 3  
US-10-078-531-7  
; Sequence 7, Application US/10078531  
; Publication No. US20030049271A1  
; GENERAL INFORMATION:  
; APPLICANT: RIOUX, STEPHANE  
; APPLICANT: BRODEUR, BERNARD R.  
; APPLICANT: HAMEL, JOSEE  
; APPLICANT: RHEAULT, PATRICK  
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING  
; TITLE OF INVENTION: DNA FRAGMENTS  
; FILE REFERENCE: PHARMA-18  
; CURRENT APPLICATION NUMBER: US/10/078,531  
; CURRENT FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: 60/269,840  
; PRIOR FILING DATE: 2001-02-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 971  
; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes  
US-10-078-531-7

Query Match 96.5%; Score 4901; DB 14; Length 971;  
Best Local Similarity 100.0%; Pred. No. 9.9e-293;  
Matches 971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 LVKEPILKQTQASSISGADYABSSGSKLKINETSGPVDVDTVDLPSDKRTTPEKIKDN 86  
DB 1 LVKEPILKQTQASSISGADYABSSGSKLKINETSGPVDVDTVDLPSDKRTTPEKIKDN 60

QY 87 LAKGPREQELKAVTENTSEKQITSGSQBSKEISLNKTVSTSNWEICDFTKGNL 146  
DB 61 LAKGPREQELKAVTENTSEKQITSGSQBSKEISLNKTVSTSNWEICDFTKGNL 120

QY 147 VGLSKSGVEKLSQTDHLVLPQADGTQLIQVASFAFTPDKTAIAEYTSRAGENGESIQ 206  
DB 121 VGLSKSGVEKLSQTDHLVLPQADGTQLIQVASFAFTPDKTAIAEYTSRAGENGESIQ 180

QY 207 LDVKGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDMKNIAEVLNPSLETISDYAF 266  
DB 181 LDVKGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDMKNIAEVLNPSLETISDYAF 240

QY 267 AHLALKQIDLPDNLKAIGELAFDNOITGKLSQMLPDGLEKIESEAPTGNP 326  
DB 241 AHLALKQIDLPDNLKAIGELAFDNOITGKLSQMLPDGLEKIESEAPTGNP 300

QY 327 KVIAGEASFQNDLSQMLPDGLEKIESEAPTGNPDDHYNRNVLLTKSGKNPSGLATEN 386  
DB 301 KVIAGEASFQNDLSQMLPDGLEKIESEAPTGNPDDHYNRNVLLTKSGKNPSGLATEN 360

QY 387 TVNVPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNLLEIPKOHNGVT 446  
DB 361 TVNVPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNLLEIPKOHNGVT 420

QY 447 ITEIGDNAFRNVDFQNKTLRKYLDEEVKLPSTIRKIGAFAPQSNLKSPEASDDLEIEKE 506  
DB 421 ITEIGDNAFRNVDFQNKTLRKYLDEEVKLPSTIRKIGAFAPQSNLKSPEASDDLEIEKE 480

QY 507 GAFMNNRIETLEKDKLVITGDAAFHNIHIAIVLPESVQIEGRSAFRONGANLIFMGS 566  
DB 481 GAFMNNRIETLEKDKLVITGDAAFHNIHIAIVLPESVQIEGRSAFRONGANLIFMGS 540

QY 567 KVKTLGEMAFLSNRLHLDLSEQKQTEIPVQAFSDNALKEVLLPASLKTIREEAFKQHL 626  
DB 541 KVKTLGEMAFLSNRLHLDLSEQKQTEIPVQAFSDNALKEVLLPASLKTIREEAFKQHL 600

Qy	627	LKQLEVASALSHIAFNALDDNDGDGBQFNKVVVVTHHNSYALADGEHFI VDPOKLSSTTV	686
Db	601	LKQLEVASALSHIAFNALDDNDGDGBQFNKVVVVTHHNSYALADGEHFI VDPOKLSSTTV	660
Qy	687	DLEKILUKLLEGLDYSYTLRQTTQTOFRDMMTAGKALLSKSNLROGEKOKFLOEAQFFLGRV	746
Db	661	DLEKILUKLLEGLDYSYTLRQTTQTOFRDMMTAGKALLSKSNLROGEKOKFLOEAQFFLGRV	720
Qy	747	DLDDKALAKAEKALVTKKATKQQLLERSINKAVLAYNNSAIKKANVVRLEKELDLLTGLV	806
Db	721	DLDDKALAKAEKALVTKKATKQQLLERSINKAVLAYNNSAIKKANVVRLEKELDLLTGLV	780
Qy	807	EKGKPLQAQATWQGVYLLKTPLEPYIYIGLVYFDKSGKLIYALDMSDTIGBGQKDAYG	866
Db	781	EKGKPLQAQATWQGVYLLKTPLEPYIYIGLVYFDKSGKLIYALDMSDTIGBGQKDAYG	840
Qy	867	NPTLVNDEONEGHALAVATLADYEGLODKITLANSKLSQLTSTRQVPTAAHYHAGIFQAI	926
Db	841	NPTLVNDEONEGHALAVATLADYEGLODKITLANSKLSQLTSTRQVPTAAHYHAGIFQAI	900
Qy	927	QNAAAAEQLLPKPGTHSEKSSSSSSANSKDRGLQSNPKTNRGHSAI LPRTGSKGSFVY	986
Db	901	QNAAAAEQLLPKPGTHSEKSSSSSSANSKDRGLQSNPKTNRGHSAI LPRTGSKGSFVY	960
Qy	987	GILGYTSVALL	997
Db	961	GILGYTSVALL	971

RESULT 4  
US-10-078-531-6  
Sequence 6, Application US/10078531  
Publication No. US20030049271A1  
GENERAL INFORMATION:  
APPLICANT: MARTIN, DENIS  
APPLICANT: RIOUX, STEPHANE  
APPLICANT: BRODEUR, BERNARD R.  
APPLICANT: HAMEL, JOSEE  
APPLICANT: RHEAULT, PATRICK  
TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING  
FILE REFERENCE PHARWA-18  
CURRENT APPLICATION NUMBER: US/10/078,531  
CURRENT FILING DATE: 2002-02-21  
PRIOR APPLICATION NUMBER: 60/269,840  
PRIOR FILING DATE: 2001-02-21  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PateatIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 971  
TYPE: PRT  
ORGANISM: Streptococcus pyogenes  
US-10-078-531-6

RESULT 5  
US-10-078-531-8  
; Sequence 8, Application US/10078531  
; Publication No. US20030049271A1  
; GENERAL INFORMATION:  
; APPLICANT: MARTIN, DENIS  
; APPLICANT: RIOUX, STEPHANE  
; APPLICANT: BRODEUR, BERNARD R.  
; APPLICANT: HAMEL, JOSEF  
; APPLICANT: RHEAULT, PATRICK  
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING  
; TITLE OF INVENTION: DNA FRAGMENTS  
; FILE REFERENCE: PHARWA-18  
; CURRENT APPLICATION NUMBER: US/10/078,531  
; CURRENT FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: 60/269,840  
; PRIOR FILING DATE: 2001-02-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1

,

```
; SEQ ID NO 8
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-8

Query Match      95.9%; Score 4871; DB 14; Length 969;
Best Local Similarity 99.6%; Pred. No. 7e-291;
Matches 965; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 28 VKPEILKQTOQASSISGADYAESSGSKLKNETSGPVDVDTVTLFSDKRTTPEKIKDNL 87
DB 1 VKPEILKQTOQASSISGADYAESSGSKLKNETSGPVDVDTVTLFSDKRTTPEKIKDNL 60
QY 88 AKGPREQELKAVTENTSESKQITSGSQLESLSLNKTVSTSNWEICDFTIKGNTLV 147
DB 61 AKGPREQELKAVTENTSESKQINSQSLESLSLNKRVSTSNWEICDFTIKGNTLV 120
QY 148 GLSKSGVEKLSQTDHLVLPQAADGTQLIOVASFAFTPDKKTAIAEYTSRAGENGESQL 207
DB 121 GLSKSGVEKLSQTDHLVLPQAADGTQLIOVASFAFTPDKKTAIAEYTSRAGENGESQL 180
QY 208 DVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNPSLETSIDYAPA 267
DB 181 DVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNPSLETSIDYAPA 240
QY 268 HLALKQIDLPDNLKAIGELAFDNOITGKLSLPROLMRLAERAFKSNHIKTIIEFRGNSLK 327
DB 241 HLALKQIDLPDNLKAIGELAFDNOITGKLSLPROLMRLAERAFKSNHIKTIIEFRGNSLK 300
QY 328 VTGEASFQDNDLSQMLPDGLEKIESEAFNGPDDHNNRVVLTGSKNPSGLATENT 387
DB 301 VTGEASFQDNDLSQMLPDGLEKIESEAFNGPDDHNNRVVLTGSKNPSGLATENT 360
QY 388 YNPNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIIPKQHNGVTI 447
DB 361 YNPNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIIPKQHNGVTI 420
QY 448 TEIGNAFNRVDFQNKTLRKYLDEEVKLPSTIRKIGAPAFQSNLKSFEASDDLEIKEG 507
DB 421 TEIGNAFNRVDFQNKTLRKYLDEEVKLPSTIRKIGAPAFQSNLKSFEASDDLEIKEG 480
QY 508 AFNNRIETLEKDKLVTTIGDAAFHNIHIAIYVLPESVQIEGRSAFRQNGANNLIFMGSK 567
DB 481 AFNNRIETLEKDKLVTTIGDAAFHNIHIAIYVLPESVQIEGRSAFRQNGANNLIFMGSK 540
QY 568 VKTLEGMAPLSNRLEHLDLSEQKQLTEIPVQAFSDNALKEVLLPASLKTIRBEAFKKNHL 627
DB 541 VKTLEGMAPLSNRLEHLDLSEQKQLTEIPVQAFSDNALKEVLLPASLKTIRBEAFKKNHL 600
QY 628 KQEVASALSHAFNALDNDGDGEQDNKVVKTHNSYALADGEHFIIVDPDKLSTIVD 687
DB 601 KQEVASALSHAFNALDNDGDGEQDNKVVKTHNSYALADGEHFIIVDPDKLSTIVD 660
QY 688 LEKILKLEGLDYSTLRQTTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVD 747
DB 661 LEKILKLEGLDYSTLRQTTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVD 720
QY 748 LDKAIAKAEKALVTKATNGQLLERSINKAVLAYNNSAIKKNVVRLEKELDLTLGLVE 807
DB 721 LDKAIAKAEKALVTKATNGQLLERSINKAVLAYNNSAIKKNVVRLEKELDLTLGLVE 780
QY 808 GKQPLAQATWVGVLTKPLPEYIIGLVNVPKSGKLIYALDMSDTIGEGQDAYGN 867
DB 781 GKQPLAQATWVGVLTKPLPEYIIGLVNVPKSGKLIYALDMSDTIGEGQDAYGN 840
QY 868 PILNVDENEGYHALAVATLADYEGLDIKTILNSKLSQLTSLRQVPTAAYHRAGIFQAIQ 927
DB 841 PILNVDENEGYHALAVATLADYEGLDIKTILNSKLSQLTSLRQVPTAAYHRAGIFQAIQ 900
QY 928 NAAAEABQLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGHSAILPRTGSKGSFVYG 987
DB 901 NAAAEABQLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGHSAILPRTGSKGSFVYG 960
```

```
QY 988 ILGYTSVAL 996
DB 961 ILGYTSVAL 969

RESULT 6
US-10-078-531-4
; Sequence 4, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 970
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-4

Query Match      95.8%; Score 4867; DB 14; Length 970;
Best Local Similarity 99.5%; Pred. No. 1.2e-290;
Matches 965; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 27 LVKPEILKQTOQASSISGADYAESSGSKLKNETSGPVDVDTVTLFSDKRTTPEKIKDNL 86
DB 1 LVKPEILKQTOQASSISGADYAESSGSKLKNETSGPVDVDTVTLFSDKRTTPEKIKDNL 60
QY 87 LAGPREBELKAVTENTSESKQITSGSQLESLSLNKTVSTSNWEICDFTIKGNTL 146
DB 61 LAGPREBELKAVTENTSESKQINSQSLESLSLNKRVSTSNWEICDFTIKGNTL 120
QY 147 VGLSKSGVEKLSQTDHLVLPQAADGTQLIOVASFAFTPDKKTAIAEYTSRAGENGESQL 206
DB 121 VGLSKSGVEKLSQTDHLVLPQAADGTQLIOVASFAFTPDKKTAIAEYTSRAGENGESQL 180
QY 207 LDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNPSLETSIDYAP 266
DB 181 LDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNPSLETSIDYAP 240
QY 267 AHLALKQIDLPDNLKAIGELAFDNOITGKLSLPROLMRLAERAFKSNHIKTIIEFRGNSL 326
DB 241 AHLALKQIDLPDNLKAIGELAFDNOITGKLSLPROLMRLAERAFKSNHIKTIIEFRGNSL 300
QY 327 KVIGEASFQDNDLSQMLPDGLEKIESEAFNGPDDHNNRVVLTGSKNPSGLATENT 386
DB 301 KVIGEASFQDNDLSQMLPDGLEKIESEAFNGPDDHNNRVVLTGSKNPSGLATENT 360
QY 387 TYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIIPKQHNGVT 446
DB 361 TYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIIPKQHNGVT 420
QY 447 ITEIGDNAPNRVDFQNKTLRKYLDEEVKLPSTIRKIGAFQSNLKSFEASDDLEIKKE 506
DB 421 ITEIGDNAPNRVDFQNKTLRKYLDEEVKLPSTIRKIGAFQSNLKSFEASDDLEIKKE 480
QY 507 GAFNNRIETLEKDKLVTTIGDAAFHNIHIAIYVLPESVQIEGRSAFRQNGANNLIFMGSK 566
DB 481 GAFNNRIETLEKDKLVTTIGDAAFHNIHIAIYVLPESVQIEGRSAFRQNGANNLIFMGSK 540
QY 567 KVKTLEGMAPLSNRLEHLDLSEQKQLTEIPVQAFSDNALKEVLLPASLKTIRBEAFKKNH 626
```

Db 541 KVKTLGEMAFSLNRLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKQNH 600  
Qy 627 LKQLEVASALSHIAFNALDDNDGDEQFNKVVVVKTHNSYALADGEHFIVDPDKLSSTIV 686  
Db 601 LKQLEVASALSHIAFNALDDNDGDEQFNKVVVVKTHNSYALADGEHFIVDPDKLSSTIV 660  
Qy 687 DLEKTLKLEGLDYSTLQTTQTTQFRDMTTAGKALLSKNLRQGEKQFLQBAQFPLGRV 746  
Db 661 DLEKTLKLEGLDYSTLQTTQTTQFRDMTTAGKALLSKNLRQGEKQFLQBAQFPLGRV 720  
Qy 747 DLDKAIAKAEKALVTKKATQNGQLLERSINKAVLAYNNSAIKAVNKRLEKELDLTLGLV 806  
Db 721 DLDKAIAKAEKALVTKKATQNGQLLERSINKAVLAYNNSAIKAVNKRLEKELDLTLGLV 780  
Qy 807 EKGGLAQATWQGVYLLKTPPLPEYVIGLVNVPFDSKGLIYALDMSDTTIGEGQKDAYG 866  
Db 781 EKGGLAQATWQGVYLLKTPPLPEYVIGLVNVPFDSKGLIYALDMSDTTIGEGQKDAYG 840  
Qy 867 NPILNVDSNEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAVHRAGIFQAI 926  
Db 841 NPILNVDSNEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAVHRAGIFQAI 900  
Qy 927 QNAAAEASQLPKPTHSEKSSSESANSKORGLQSNPKTNRGRHSAILPRTGSKGSFVY 986  
Db 901 QNAAAEASQLPKPTHSEKSSSESANSKORGLQSNPKTNRGRHSAILPRTGSKGSFVY 960  
Qy 987 GILGYTSVAL 996  
Db 961 GILGYTSVAL 970

## RESULT 7

US-10-078-531-5  
; Sequence 5, Application US/10078531  
; Publication No. US20030049271A1  
; GENERAL INFORMATION:  
; APPLICANT: MARTIN, DENIS  
; APPLICANT: RIOUX, STEPHANE  
; APPLICANT: BRODEUR, BERNARD R.  
; APPLICANT: HAMEL, JOSEE  
; APPLICANT: RHEAULT, PATRICK  
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING  
; FILE REFERENCE: PHARMA-18  
; CURRENT APPLICATION NUMBER: US/10/078,531  
; CURRENT FILING DATE: 2002-02-21  
; PRIOR FILING DATE: 2001-02-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 963  
; TYPE: PRN  
; ORGANISM: Streptococcus pyogenes  
US-10-078-531-5

## Query Match

Best Local Similarity 95.1%; Score 4831; DB 14; Length 963;  
Matches 957; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
Qy 34 KOTQASSISGADYAESGSKLKTNETSGPVDVTDVTLFSDKRTTPKIKDNLAKGPKE 93  
Db 1 KOTQASSISGADYAESGSKLKTNETSGPVDVTDVTLFSDKRTTPKIKDNLAKGPKE 60  
Qy 94 QELKAVTENTESEKQITSGSQLEQSKESLSLNKTVFSTSNWEICDFITKNTLVLSKSG 153  
Db 61 QELKAVTENTESEKQINSGSQLEQSKESLSLNKTVFSTSNWEICDFITKNTLVLSKSG 120  
Qy 154 VEKLSQTDHLVLPQAAQDTQLIQVAFPTPKKTAIAEYTSRAGENGESQLDVGKE 213  
Db 121 VEKLSQTDHLVLPQAAQDTQLIQVAFPTPKKTAIAEYTSRAGENGESQLDVGKE 180  
Qy 214 IINEGEVFNYSLLKKVTIPTGYKHIGQDAFVNDKNKNAEVLNLPESLETISDYAFALHAKQ 273

Db 181 IINEGEVFNYSLLKKVTIPTGYKHIGQDAFVNDKNKNAEVLNLPESLETISDYAFALHAKQ 240  
Qy 274 IDLPDNLKAIGELAFFNQITGKLSPLQRLMRLAERAFKSNHIKTIIEFRGNSLKVIGEAS 333  
Db 241 IDLPDNLKAIGELAFFNQITGKLSPLQRLMRLAERAFKSNHIKTIIEFRGNSLKVIGEAS 300  
Qy 334 FQDNDLSQMLPDGLEKIESEAFPTGNDDHYNRVVLTWSKGNPSGLATENTYVNPDK 393  
Db 301 FQDNDLSQMLPDGLEKIESEAFPTGNDDHYNRVVLTWSKGNPSGLATENTYVNPDK 360  
Qy 394 SLWQESPEIDYTWLWLEBDFYQKNSVTGFGNKGLQKVKRNKNLEIPKOHNGVTITTEIGDN 453  
Db 361 SLWQESPEIDYTWLWLEBDFYQKNSVTGFGSKGLQKVKRNKNLEIPKOHNGVTITTEIGDN 420  
Qy 454 AFRNVDFONKTLRKYLEEVKLPSTIRKIGAFQSNLKSFEASDDLEIEKEGAFMNNR 513  
Db 421 AFRNVDFONKTLRKYLEEVKLPSTIRKIGAFQSNLKSFEASDDLEIEKEGAFMNNR 480  
Qy 514 IETLELKOKLVTIGDAAFHINHIYAIIVLPSVQIEGRSAFRONCANNLIIFMGSKVKTLGE 573  
Db 481 IETLELKOKLVTIGDAAFHINHIYAIIVLPSVQIEGRSAFRONCANNLIIFMGSKVKTLGE 540  
Qy 574 MAFLSNLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKQNHKQLEVA 633  
Db 541 MAFLSNLEHLDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKQNHKQLEVA 600  
Qy 634 SALSHIAPNALDDNDGDEQFNKVVVVKTHNSYALADGEHFIVDPDKLSSTIVLEKILK 693  
Db 601 SALSHIAPNALDDNDGDEQFNKVVVVKTHNSYALADGEHFIVDPDKLSSTIVLEKILK 660  
Qy 694 LIEGLDYSTLRQTTQTTQFRDMTTAGKALLSKNLRQGEKQFLQBAQFPLGRVLDKAI 753  
Db 661 LIEGLDYSTLRQTTQTTQFRDMTTAGKALLSKNLRQGEKQFLQBAQFPLGRVLDKAI 720  
Qy 754 KAERKALVTKKATNGQLLERSINKAVLAYNNSAIKAVNKRLEKELDLTLGLVEGKGLA 813  
Db 721 KAERKALVTKKATNGQLLERSINKAVLAYNNSAIKAVNKRLEKELDLTLGLVEGKGLA 780  
Qy 814 QATWQGVYLLKTPPLPEYVIGLVNVPFDSKGLIYALDMSDTTIGEGQKDAYGNPILNVD 873  
Db 781 QATWQGVYLLKTPPLPEYVIGLVNVPFDSKGLIYALDMSDTTIGEGQKDAYGNPILNVD 840  
Qy 874 EDNEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAVHRAGIFQAIQNAAAEA 933  
Db 841 EDNEGHALAVATLADYEGLDIKTILNSKLSQTSIROVPTAAVHRAGIFQAIQNAAAEA 900  
Qy 934 EQLLPKPTHSEKSSSESANSKORGLQSNPKTNRGRHSAILPRTGSKGSFVYGILGYTS 993  
Db 901 EQLLPKPTHSEKSSSESANSKORGLQSNPKTNRGRHSAILPRTGSKGSFVYGILGYTS 960  
Qy 994 VAL 996  
Db 961 VAL 963

## RESULT 8

US-10-078-531-3  
; Sequence 3, Application US/10078531  
; Publication No. US20030049271A1  
; GENERAL INFORMATION:  
; APPLICANT: MARTIN, DENIS  
; APPLICANT: RIOUX, STEPHANE  
; APPLICANT: BRODEUR, BERNARD R.  
; APPLICANT: HAMEL, JOSEE  
; APPLICANT: RHEAULT, PATRICK  
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING  
; FILE REFERENCE: PHARMA-18  
; CURRENT APPLICATION NUMBER: US/10/078,531  
; CURRENT FILING DATE: 2002-02-21  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR FILING DATE: 2001-02-21

; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 951  
; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes  
US-10-078-531-3

Query Match 94.0%; Score 4775; DB 14; Length 951;  
Best Local Similarity 99.3%; Pred. No. 5.5e-285;  
Matches 944; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
  
QY 46 DYAESGSKLKNETSGPVDVTVDLFSDKRTTPEKIKONLAKGPREQELKAVTENTES 105  
DB 1 DYAESGSKLKNETSGPVDVTVDLFSDKRTTPEKIKONLAKGPREQELKAVTENTES 60  
  
QY 106 EKQITSGSOLEKSKSLINKTPVSTSNWEICDFITKGNLTVLGSKSGVEKLSQTDHLVL 165  
DB 61 EKQITSGSOLEKSKSLINKTPVSTSNWEICDFITKGNLTVLGSKSGVEKLSQTDHLVL 120  
  
QY 166 PQQAADGTOLIQVAFAPDPDKTAIAEYTSRAGENGEISQDVGKEIINEGEVFNLSYL 225  
DB 121 PQQAADGTOLIQVAFAPDPDKTAIAEYTSRAGENGEISQDVGKEIINEGEVFNLSYL 180  
  
QY 226 LKXVTIPTGYKHIGQDAFVDNKNIAEVNLPESLETISDYAFALHALKQIDLPNLKAIGE 285  
DB 181 LKXVTIPTGYKHIGQDAFVDNKNIAEVNLPESLETISDYAFALHALKQIDLPNLKAIGE 240  
  
QY 286 LAFPDNQITGKLSLPROMLRAERAFKSNHIKTIIEPRGNSLKVIGEASFQDNDLSQMLP 345  
DB 241 LAFPDNQITGKLSLPROMLRAERAFKSNHIKTIIEPRGNSLKVIGEASFQDNDLSQMLP 300  
  
QY 346 DGLEKIESEAFNGPDDHNNRVLVLTWSKGNPGLATENTYVNPDKSLWQESPEIDYT 405  
DB 301 DGLEKIESEAFNGPDDHNNRVLVLTWSKGNPGLATENTYVNPDKSLWQESPEIDYT 360  
  
QY 406 KWLDEDFTYQKNSVTGFSNGKIQKVRKNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTL 465  
DB 361 KWLDEDFTYQKNSVTGFSNGKIQKVRKNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTL 420  
  
QY 466 RYVDLEEVKLPSTIRKIGAFQSNLKSFEASDDLEEIKEGAFMNNRIETLELQKLVY 525  
DB 421 RYVDLEEVKLPSTIRKIGAFQSNLKSFEASDDLEEIKEGAFMNNRIETLELQKLVY 480  
  
QY 526 IGDAAFHINHIIYAI VLPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFISNRLHL 585  
DB 481 IGDAAFHINHIIYAI VLPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFISNRLHL 540  
  
QY 586 LSEQKQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSALSHIAFNALD 645  
DB 541 LSEQKQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSALSHIAFNALD 600  
  
QY 646 DNDGDEQFNKVVVTHNSYALADGEHPI VDPDKLSSTIVLEKILKUIEGLDYSTLRQ 705  
DB 601 DNDGDEQFNKVVVTHNSYALADGEHPI VDPDKLSSTIVLEKILKUIEGLDYSTLRQ 660  
  
QY 706 TTQTFQFRDMMTAGKALLSKSNLROCKQFLOEAPFLGRVLDLKAIAKAEKALYTKKAT 765  
DB 661 TTQTFQFRDMMTAGKALLSKSNLROCKQFLOEAPFLGRVLDLKAIAKAEKALYTKKAT 720  
  
QY 766 KNGQLLERSINKAVLAYNNSAIKKANVRKLEKLELOLLTGLVEGKPLAQATVWQGVYLLK 825  
DB 721 KNGQLLERSINKAVLAYNNSAIKKANVRKLEKLELOLLTGLVEGKPLAQATVWQGVYLLK 780  
  
QY 826 TPLPLPEYIYGLNVVPDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDEEDNRYHALAVA 885  
DB 781 TPLPLPEYIYGLNVVPDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDEEDNRYHALAVA 840  
  
QY 886 TLADYEGLDIKTILNSQLSITSIQVPTAAVHRAGI FOAIONAAEAEOQLPKPGTHSE 945  
DB 841 TLADYEGLDIKTILNSQLSITSIQVPTAAVHRAGI FOAIONAAEAEOQLPKPGTHSE 900  
  
QY 946 KSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKGSFVYGLGYTSVAL 996

DB 901 KSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKGSFVYGLGYTSVAL 951  
  
RESULT 9  
US-10-091-007-24  
; Sequence 24, Application US/10091007  
; Publication No. US20030170782A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics limited  
; APPLICANT: Le Page, Richard W F Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PWC/P21978WO  
; CURRENT APPLICATION NUMBER: US/10/091,007  
; CURRENT FILING DATE: 2002-03-06  
; PRIOR APPLICATION NUMBER: GB 9921125.2  
; PRIOR FILING DATE: 1999-09-07  
; NUMBER OF SEQ ID NOS: 276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 24  
; LENGTH: 1055  
; TYPE: PRT  
; ORGANISM: Streptococcus agalactiae  
US-10-091-007-24  
  
Query Match 74.4%; Score 3777.5; DB 14; Length 1055;  
Best Local Similarity 71.7%; Pred. No. 1.5e-223;  
Matches 756; Conservative 105; Mismatches 140; Indels 53; Gaps 5;  
  
QY 2 KKHKTVALTLTTSVVVTHNQEVSLVKEPILKQTPASSISIGADYAESGSKSLKINET 61  
DB 3 KKHKTVALTLTTSVVVTHNQEVSLVKEPILKQTPASSISIGADYAESGSKSLKINET 61  
  
QY 62 SGPVDDTVDLFSDKRTTPEKIKONLAKGPREQELKAVTENT-ESEKQITSSQLEQSK 120  
DB 62 NSTVDETSDLFSDGNSNNSKTESVVSVDPKVPEKAPKEVTQEASNSNDASKVEVPKQ 121  
  
QY 121 SLSLKNKTPVSTSNWEICDFITKGNLTVLGSKSGVEKLSQTDHLVLPQADGTOLIQVAS 180  
DB 122 DTASKKETLETSWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLPQADGTOLIQVAS 181  
  
QY 181 FAFTPDKKTAAEYTSRAGENGEISQDVGKEIINEGEVFNLSYLLKVTIPTGYKHIGQ 240  
DB 182 FAFTPDKKTAAEYTSRAGENGEISQDVGKEIINEGEVFNLSYLLKVTIPTGYKHIGQ 241  
  
QY 241 DAFVONKNIAEVNLPESLETISDYAFALHALKQIDLPNLKAIGELAFPDNQITGKLSLP 300  
DB 242 DAFVONKNIAEVNLPESLETISDYAFALHALKQIDLPNLKAIGELAFPDNQITGKLSLP 301  
  
QY 301 RQMLAERAFKSNHIKTIIEPRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEAFNGP 360  
DB 302 RHLIKLAERAFKSNHIKTIIEPRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEAFNGP 361  
  
QY 361 GDDYHNNRVLVLTWSKGNPGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420  
DB 362 GDEYHNNRVLVLTWSKGNPGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 421  
  
QY 421 GFSNKGLOKVRKNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480  
DB 422 GFSNKGLOKVRKNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 481  
  
QY 481 KIGAFAFOSNLLKSFASDDLEEIKEGAFMNNRIETLELQKLVITIGDAAFHINHIIYAI 540  
DB 482 KIGAFAFOSNLLKSFASDDLEEIKEGAFMNNRIETLELQKLVITIGDAAFHINHIIYAI 541  
  
QY 541 LPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFISNRLHLDLSEKQLETPVQAF 600  
DB 542 LPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFISNRLHLDLSEKQLETPVQAF 601  
  
QY 601 SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSALSHIAFNALDNDGDEQFNKVVV 660  
DB 602 SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSALSHIAFNALDNDGDEQFNKVVV 661



```
QY 661 THNSYALADGHRFIVDPDKLSTIVDLKILKLEGLDYSTLRQTTOTQFRDMWTAGKA 720
DB 662 THNSHMLADGHRFIVDPDKLSTIVDLKILKLEGLDYSTLRQTTOTQFRDMWTAGKA 721
QY 721 LLSKNLRQGEKQKFLQBAQFPLGRVDLDKAKAEKALVTKATKNGOLLERSINKAVL 780
DB 722 LLSKNLRQGEKQKFLQBAQFPLGRVDLDKAKAEKALVTKATKNGOLLERSINKAVL 781
QY 781 AYNSAIKANKVRLKLELDLTLGLVEGKPLAQATMWQGVYLLKTPPLPEYIIGLVNY 840
DB 782 AYNSAIKANKVRLKLELDLTLGLVEGKPLAQATMWQGVYLLKTPPLPEYIIGLVNY 841
QY 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEBNEGYHALAVATLADYGLDKITLN 900
DB 842 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEBNEGYHALAVATLADYGLDKITLN 901
QY 901 SKLSOLTSIROVPTAAVHRAGIFQAIQNAAAEQLLPKPGTHSKSSSSSANSKDRGL 960
DB 902 SBLDKIKAIRQIPLAKYHRLGIFQAIRNAAEADRLPK-----TPKGYLNEVPNVRKKQM 957
QY 961 QSNPK-----TN-----RGRHSA 973
DB 958 EKNLKPVDYKTPFNKALPNEKVDGDRAAKGHNINAEITNNSVAVTPIRSEQQLHKSQSDV 1017
QY 974 ILPRTSGKSFYVIGILYTSVALLSLITAIKKK 1007
DB 1018 NLPQTSSKNNFIYBILGYVSLCLLFLVTAGKKG 1051

RESULT 10
US-10-476-614-2
; Sequence 2, Application US/10476614
; Publication No. US2004017113A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem Inc.
; TITLE OF INVENTION: Antigens of Group B Streptococcus and Corresponding DNA Fragments
; FILE REFERENCE: 74872-81
; CURRENT APPLICATION NUMBER: US/10/476,614
; CURRENT FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: US 60/287,712
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Streptococcus
US-10-476-614-2

Query Match 74.4%; Score 3777.5; DB 16; Length 1055;
Best Local Similarity 71.7%; Pred. No. 1.5e-223;
Matches 756; Conservative 105; Mismatches 140; Indels 53; Gaps 5;

QY 2 KGHKLTVALTLTVSNVTHNQEVPSIVKEPIIKQTKQASSISGADYABSSGSKLKNET 61
DB 3 KGHKLTVALTLTVSNVTHNQEVPSIVKEPIIKQTKQASSISGADYABSSGSKLKNET 61
QY 62 SGVPDDTVTDLFSKRTTPEKIKDLNLAGPREQELKAVTENT-ESEKQITSGSQLEQSK 120
DB 62 NSTVETVSDLFSGNSNNSSKTSVSDPKQVPKAPVETQEAASNSNDASKVEVPKQ 121
QY 121 SUSLNTVPSTNWEICDPITKGNLTVGLSGKGVKLSQTDHLVLPQAAADTQLIQVAS 180
DB 122 DTASKKETLETSTWEAKDPVTRGDTLVGFSKSGINKLSQTSHLVLPQAAADTQLTQVAS 181
QY 181 FAPTPDKKTAIAEYTSRAGECEISOLDVDGKEIINEGEVENSLLKVKVTIPTGKHIGQ 240
DB 182 FAPTPDKKTAIAEYTSRAGECEISOLDVDGKEIINEGEVENSLLKVKVTIPTGKHIGQ 241
QY 241 DAFVDMKNIAEVLNPSLETISDYAFALHAKQIDLPNLKAIAGELAFPDNQITGKLSIP 300
DB 242 DAFVDMKNIAEVLNPSLETISDYAFALHAKQIDLPNLKAIAGELAFPDNQITGKLSIP 301
```

```
QY 301 RQMLRLAERAPKSNHIIKTIEPRGNSLKVIGBASQDNDLSQMLPDGLEKIESEAFQNP 360
DB 302 RHLIKLAERAPKSNRIQTVEFLGSKLVIGBASQDNNLRNMLPDGLEKIESEAFQNP 361
QY 361 GDDHNNRVVLWTKSGKNPSGLATENTVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB 362 GDEHNNQVLUWTKSGKNPSGLATENTVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 421
QY 421 GFSNGKQKQVKNRNKLEIPKQHNGVTITEIGDANFRNVDFQNKTLRKYDLEEVKLPSTIR 480
DB 422 GFSNGKQKQVKNRNKLEIPKQHNGVTITEIGDANFRNVDFQNKTLRKYDLEEVKLPSTIR 481
QY 481 KIGAFAPQSNLKFSEASDDLEEBEKEGAFMNNRIETLEKOKLVITIGDAAAPHINHIYAV 540
DB 482 KIGAFAPQSNLKFSEASDDLEEBEKEGAFMNNRIETLEKOKLVITIGDAAAPHINHIYAV 541
QY 541 LPESVQEIIGRSAPFQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDLSQKOLTEIPVOAF 600
DB 542 LPESVQEIIGRSAPFQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDLSQKOLTEIPVOAF 601
QY 601 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSASHIAFNALDDNDGDQFQNKVVVK 660
DB 602 SDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSASHIAFNALDDNDGDQFQNKVVVK 661
QY 661 THNSYALADGHRFIVDPDKLSTIVDLKILKLEGLDYSTLRQTTOTQFRDMWTAGKA 720
DB 662 THNSHMLADGHRFIVDPDKLSTIVDLKILKLEGLDYSTLRQTTOTQFRDMWTAGKA 721
QY 721 LLSKNLRQGEKQKFLQBAQFPLGRVDLDKAKAEKALVTKATKNGOLLERSINKAVL 780
DB 722 LLSKNLRQGEKQKFLQBAQFPLGRVDLDKAKAEKALVTKATKNGOLLERSINKAVL 781
QY 781 AYNSAIKANKVRLKLELDLTLGLVEGKPLAQATMWQGVYLLKTPPLPEYIIGLVNY 840
DB 782 AYNSAIKANKVRLKLELDLTLGLVEGKPLAQATMWQGVYLLKTPPLPEYIIGLVNY 841
QY 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEBNEGYHALAVATLADYGLDKITLN 900
DB 842 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEBNEGYHALAVATLADYGLDKITLN 901
QY 901 SKLSOLTSIROVPTAAVHRAGIFQAIQNAAAEQLLPKPGTHSKSSSSSANSKDRGL 960
DB 902 SBLDKIKAIRQIPLAKYHRLGIFQAIRNAAEADRLPK-----TPKGYLNEVPNVRKKQM 957
QY 961 QSNPK-----TN-----RGRHSA 973
DB 958 EKNLKPVDYKTPFNKALPNEKVDGDRAAKGHNINAEITNNSVAVTPIRSEQQLHKSQSDV 1017
QY 974 ILPRTSGKSFYVIGILYTSVALLSLITAIKKK 1007
DB 1018 NLPQTSSKNNFIYBILGYVSLCLLFLVTAGKKG 1051

RESULT 11
US-10-369-493-11351
; Sequence 11351, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11351
```

```
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Methanosarcina mazei
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(669)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-11351

Query Match
Best Local Similarity 4.8%; Score 241.5; DB 15; Length 669;
Matches 124; Conservative 84; Mismatches 214; Indels 125; Gaps 28;

QY 136 ICDFTKGNLTGLSKSGVEKLSQTDHLVLPQQAADGTOLIQVASFAFTPDKKAIAEY 195
Db 177 LCMSTTHGAATVSSEQTLPXRLT--LTVGLTSDGTS-VTVTGVSG-SPDSVPIP--- 229

QY 196 SRAGENGEISQLDVDGKEIINEGEVFNLSYL--LKKVTPTGKYGKIGQDAFVNKNAE 253
Db 230 -----GEI-----DGLPVTTIGASACKALTSLTSPDSTTIDASAFYGTNLSVT 278

QY 254 LPSELTISDYAF-AHLALKQIDLDPNKAIGELAFEDNQITGKLS--LPROLMRLAERA 310
Db 279 MPDSVTITIGSSAFYCGTGLASVTIIPNSVTIIGSSAFY--CTGLISVTIPNSVTIIGSSA 336

QY 311 FKS-NHIKTIIEPRGNSLVIGE-----ASFQDNDLSQLMLPDGLEKIES 353
Db 337 FYGCTNLVTMMENGNAPTTVGSNWASGTVNLVAYISEGATNFTTPENNEVPCYPALTAATA 396

QY 354 E-----AFTG---NPGDDHYNRVVLWTKSGKNPS-GLATENTY----- 388
Db 397 EPTSNIVYGSAPLTVKFTVTGVCANALDWYFNDSTV-DSTARNPSTYDTPTGYSIKLN 455

QY 389 --VNPDKSLWQESPEID--YTKWLE--EDFTYQKN-----SVTGFSGKGLQKVENK 439
Db 456 VSNP-----WGSDEKVTAYITVAEPVDNFTYLSGDTSVTITGSGPG-----GDV 504

QY 440 KOHNGVTITEIGDGNFRNVDQNKTLRKVDLEEVKLPSTIRKIGAFAPQS--NNLKSFEAS 498
Db 505 STIGNLPVTAISDVFKS-----WTNITSVPIPSVTIIGSSAFYCGTGLTSVIP 555

QY 499 DLEIKGGAFFN-NRIETLEKDKLVITIGDAAFH--INHIAVLVPESVQIEGRSAFRQN 556
Db 556 NSVTITIGNAFNGCTGLTSMITGNVTKIDRMFGCSALISWIPDSVTIIGDGAFFYQ- 614

QY 557 GANNLIFMGSKVKTIGEMAFNLRLHLDLSEKQQLTEIPVOAFSD-NALKEVLLPASLK 615
Db 615 -CSNL-----ASVTIPDSVTITIDSYAFRDCNALTSVTIPDSVT 651

QY 616 TIREBAF 622
Db 652 NVESRLF 658
```

```
RESULT 12
US-10-369-493-11352
; Sequence 11352, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11352
```

```
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Methanosarcina mazei
US-10-369-493-11352

Query Match
Best Local Similarity 4.7%; Score 237.5; DB 15; Length 668;
Matches 128; Conservative 94; Mismatches 207; Indels 211; Gaps 20;

QY 13 TTVSVVTHNQEVFSIAVKSPILKQTQAS-----SSISGADYAESSCKSKLNETSGPVDD 67
Db 147 TSDIGVLYNKDVTTLIRCPGKGTGSAIPNSVTIISVFPDSCGLTITIGSGVTSIDN 206

QY 68 TVTDLFSKRTTPKIKONLAKGPREQELKAVTENTESEKQITSSQLEQSKESLSLNTK 127
Db 207 SV---FSDC-----SALTAI-----DVDNDNSMYASIDGVLYNK- 237

QY 128 VPSTSNWEICDPITKGNLTGLSKSGVEKLSQTDHLVLPQQAADGTOLIQVASFAFTPD 187
Db 238 -----DITTLIQCP-----DITTLIQCP----- 247

QY 188 KTAIAEYTSRAGENCEISQLDVDGKEIINEGEVFNLSYLLKKVTPTGKYGKIGQDAFVN 247
Db 248 -----GKTGSVT--IPDSVTIIGDGAFFYCGTALTSTVTIIPDSVTIIGTSAFQGC 294

QY 248 NIAEVLNPESLETISDYAFAH-LALKQIDLDPNKAIGELAFEDNQITGKLSLRLQML 306
Db 295 SITSNIGNNVTSIGNSAFCDIALTSVTIGNNVISIGSYAFQGCALTSVTIIPDSVTKI 354

QY 307 AERAPKSNHIKTIIEPRGNSLVIGESAF-QDNDLSQLMLPD----- 346
Db 355 GSSAFQGCISITSVNVIGSVTSIGNSAFQCSALTFTVTIIPDSVTIILDYAFQDCSRLT 414

QY 347 ---GLEKISEAFNGDGHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEI 402
Db 415 TIGSVTNIDSAFSG-----CSALTAID--VDNDNSIYASIDGV 452

QY 403 DYTQWLEEDFTYQKNSTVGFSGNKGKQVKRNKNLEIPKQHNGVTITEIGDGNFRNVD-- 459
Db 453 LYNK-----DITTLIQCPG-----KTGSVTIPD-----SVTIGDSAFRDCSGLT 493

QY 460 -----FQNKTLRKVDLEEVKLPSTIRKIGAFAPQS--NNLKSFEASDDLEE 503
Db 494 IVPIGNNVTSIGSSAFYCGTV---LTSVPIGNNVTSIGSSAFYCGTSLTSVTIGNNVTS 549

QY 504 IKEGAFNN-NRIETLEKDKLVITIGDAAFH-----INHIA 538
Db 550 ISSAFQCSALTFTVTIIPDSVTIIGDIAFSDCSGLTFTVTIGNNVTSIGSSAFYGTALT 609

QY 539 IVPESVQIEGRSAFRONGANNLIFMGSKVKVTLGEMAFLS 578
Db 610 IAPDSVTISGSAFRDCTNLTSTVTIGNNVTSIGSSAFRS 649
```

```
RESULT 13
US-10-617-320-2948
; Sequence 2948, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
```



175 LEYIGTSAFSFOKLLKLTFTSSSKLEL-----ISHFAFANLSNLEKLTLPKS 222  
 169 AAD-GTQLIQVA-----SFA-----FTDPKKTAAIEYTSRAGENGESISQLD 208  
 223 VKTGLSNLPLRTTSLKHVDVEEGNESFASVDGVLFSDK-KTQLIYYPSEQ--KNDESXYKTP 279  
 209 VDGKEIINEGEVFNYSYLLKVTIPTGYKHHGODAFVFNKNAEAVNLPESLETISDYAF-A 267  
 280 KETKELASYSFNKNSY-LKKLELNEGLEKI GTFAFADAIKLEEISLPSNLETIERLAFY 338  
 268 HLAKOIDLPDLNKAIGELAF-----FDNQITGKLSLPROLMRLAERAFKSNHIK 317  
 339 NLEKLELIDPNVQNFQKHVWGLPKLSLTIGNNIN---SLPFFFLSGVLDLSLKEIHK 395  
 318 -----TIFRGNLS--LKVIGEASFQ-DNDLSQMLPDGLEKIESEAPT 357  
 396 NKSTEPSVKKDTFPAIPETVKFYVTSHEIKVDLKSNLSTSDI-----IVEKVDNIKQETDV 451  
 358 GNPGDHYNRRVVLWTK-----SGKNPSSLATENTVVPDKSLW---QESPEIDY 404  
 452 AKP-KKNSGGVGVWVKDGLWYLYNBSGSMATGWV-----KDKGLWYLYNBSGSM-A 502  
 405 TKWLESD--PTYQKNS---VTGF-SNKGL 427  
 503 TGWVKDKGLWYLYNBSGSMATGWVKDKGL 531  
 RESULT 16  
 US-10-472-928-4474  
 ; Sequence 4474, Application US/10472928  
 ; Publication No. US20050020813A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHIRON Spa  
 ; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS  
 ; FILE REFERENCE: P026926W0  
 ; CURRENT APPLICATION NUMBER: US/10/472,928  
 ; CURRENT FILING DATE: 2003-09-26  
 ; PRIOR APPLICATION NUMBER: GB-0107658.7  
 ; PRIOR FILING DATE: 2001-03-27  
 ; NUMBER OF SEQ ID NOS: 4979  
 ; SOFTWARE: SeqWin99, version 1.03  
 ; SEQ ID NO 4474  
 ; LENGTH: 621  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus pneumoniae  
 ; FEATURE:  
 ; OTHER INFORMATION: choline binding protein PcpA (pcpA)  
 ; OTHER INFORMATION: Cellular location: outside  
 ; OTHER INFORMATION: Feature of note: WY motif  
 ; OTHER INFORMATION: Similar to strain R6 sequence 15903986 (O.E+01)  
 US-10-472-928-4474  
 Query Match 4.2%; Score 213.5; DB 17; Length 621;  
 Best Local Similarity 23.2%; Pred. No. 0.00014;  
 Matches 132; Conservative 75; Mismatches 175; Indels 187; Gaps 31  
 QY 6 KTVALLTFTVSVV-----THNOEVF-----SLVKEPILK-----Q 35  
 DB 3 KTWILSLTTAAVILAAVYVNEPILADTPSSEVIKETKVGSIIQQNNIKYKVLTVEGNIGT 62  
 QY 36 TOASSISFGADY-AESSGK-----SKLKINETSPPDVTVDLFS-----DK 76  
 DB 63 VQVNGVTVPEFSGAQDGKPTPTKLTVDGKVTVTVEVASQAFSYPDETGRIVYPPSS 122  
 QY 77 RTTPEKIDKNLAKGPRQELKAVTENTSEBKQITSGSQLESK----- 120  
 DB 123 ITTPSSIKTKKGFGHGSKAKTII-----FDKGSQLEKIEDRAPDSELEIELPAS 174  
 QY 121 -----SLSLNK-TVPSTSNWEICDFITKGNLTVLGUSKGVSKLQTDHVLVPSQ 168  
 DB 175 LEYIGTSAFSFOKLLKLTFTSSSKLEL-----ISHFAFANLSNLEKLTLPKS 222







```
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7646
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Candida albicans
; NAME/KEY: MISC FEATURE
; LOCATION: (1881)..(1881)
; OTHER INFORMATION: X-any amino acid
US-10-032-585-7646

Query Match      4.2%; Score 211; DB 14; Length 1881;
Best Local Similarity 19.9%; Pred. No. 0.00094;
Matches 227; Conservative 173; Mismatches 431; Indels 310; Gaps 48;

QY 5 LKTVALLTTSVVTNN-----QEVSLVKEPILKQTQASSIS 43
Db 689 LPTVLTVDVYVNLVDNFIRIKRALFHPDSSPRGKISYELFEELDTKVLNLTKEIQ--T 746

QY 44 GADYAESGSKSLKINETSPPVDDTVTDLFSDKRTTPEKIDNLAGPREQELKAVTENT 103
Db 747 EKENAESNDK---ELNEKI-----EKLNLSTKLET---KLED-----KEQELAKIQEDH 790

QY 104 EG--EKQITGSQLSQSKESLSINKTVPSTNSWEICDFTTKNTLVGLSKSGVEKLSOTD 161
Db 791 KSLNEKFLVTANSLOGIKARTKESITSGPDQQLQELQALKKGTSTSLKQLKEXKLDSTGE 850

QY 162 -----HLVPSQAADGTOLIQ-----VASPFTPDKTAIAEYTSR 197
Db 851 QAKKKLEGGINNMTDRDLFHLKXKSEAE--TQIKQREPERPNLTYSFENTKQDYEIQINN 909

QY 198 AGENGEISO-----LDVDGK-----EIIINEGEVENSYLLKKVTIPTGYKHIG 239
Db 910 KNSNNEFKOKINELSKKTESLTDNFKNAKQLEKLRDTEENNEHMLDKLRAS-----VA 965

QY 240 QDAFVDNKNIAE---VNLPSLETISDYAFALHAKQIDLPNL--KAIGELAFFNQITG 295
Db 966 YNDLKKAKGSESEETVKAKELETITSKI-----DNLEKELKQSKKKELEG 1013

QY 296 KLSLPQRLMLARAPK--SNIKIETEPGNSLVKIGVASFOODLSQLMLPDGLEKIES 353
Db 1014 QL---QNTDSTNEXFKELEDELKSIK-----KSNKEISSQNSBLIQ-----KLEKTEK 1059

QY 354 EAPTGNPGGDHNNRVLVTKSGKNPSGLATENTVYVDPKSLWQSPEDIDYTKWLEEDFT 413
Db 1060 DL---QAKDEIDKLKAEFK-----SNIDNLSISSLSQSKLKEA-----EESH 1101

QY 414 YQKNSVTGFSNKGLOKVK-----RNKNLETPKHQNGVTITEIGDAPRNVDPNQKTLRYD 469
Db 1102 STKDEHSSLS--ENLKKLKEVENTKTSMAK-----LSAKIEBKKAQDEIETKTKHITD 1155

QY 470 LSEVKLPSTIRKIGAFAPSNLKS-----PEASDLEEI-----KEGAPNNRIETLEL 519
Db 1156 LQE-----EHAQKQSPESRNDIKNLDKANKEKLSNLEKEKTELNNKLTQORE 1211

QY 520 KDKLVITGDAAPHINHIYAIVLPESVQEIGRSFQNGANNLI FMGSKVKTGEMAFLSN 579
Db 1212 K---ISDLETSVAISEDKSKSLKHIEDLKR-----EKIKL---ETTLKEN 1251

QY 580 RLEHLDLSQKQJTEIPVOAFSDNALKEVLLPASLKTIIE-----EAFKKNH 626
Db 1252 EETMPEKKEQOLVV-----NDCKE--LEACKLKTETKEINLDIRKLEAKSDH 1301

QY 627 LKQLEVASALSHAFNALDNDGDGQFNKVVVKTHNSYAL-----668
Db 1302 DTERKKLSL-----IEDTKSESE---KNVKNLEKLEKGEREVRDIQSLAAK 1351

QY 669 -ADGSHFVDPDKLSSTIVDLKILKIEGLDYSTLRQTTQT-----QFRDMWTTA 717
```

```
Db 1352 TTDWEKIKTTLDKVLKEKSDLEKTNK--ESVD--TLKKEVENLAKKEISLLEBQKDDTTK 1407
QY 718 GKALLSKSNLRQGBKQKFLQEAQFPLGRVDLDKAIKAEKALVTYKATK---NQOLLERS 774
Db 1408 YKELAAQLE---TKTSNLDSTTMELEKTELELKKVRNELTEATSELTKLQDNNQSLTEE 1463
QY 775 INKAVLAYNNS-----AIK--KANVKRL-----EKELD 800
Db 1464 LEKKAALTKSSKDLVCGNQSSELODLSKSVKSELKNFENKYNQETSLSKDEIEEKQKE 1523
QY 801 LLTGLVBEKGPLAQATWVGYYLLKTPLPPEYTYIGLVNVPYDKGKLIYALDMSDTIGEG 860
Db 1524 IVTLQTELKORISEVEKERAMSENSETVIVEKSDKISLESKINSIKENHSKITTHNE 1583
QY 861 QKDAYGNPILNVDEDNEG-----YHALAVATLADYEGLDIKITILN 900
Db 1584 QKTSLKQDIAKLSQDHSQAQTLQEDKENQLKELKASLEKHNTEGATSIETEEKNQIKELSE 1643
QY 901 SKLSQLSIRQVPTAAVHRAGIFQAIQNAAEASOLLPKPGTHSEK--SSSESANSKDRG 959
Db 1644 TIKSLTELKTSGDALKQSQKKEYTKLTGNSDTSKLEKQLEBELEKVKSDLQTADEKLG 1703
QY 960 L 960
Db 1704 I 1704

RESULT 20
US-10-437-963-135452
; Sequence 135452, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbaszuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 135452
; LENGTH: 2295
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURES:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37127C.1.pep
US-10-437-963-135452

Query Match      4.1%; Score 210; DB 16; Length 2295;
Best Local Similarity 20.1%; Pred. No. 0.0014;
Matches 215; Conservative 176; Mismatches 411; Indels 270; Gaps 47;

QY 16 SWVTHNQEVSLVKEPILKQTQASSISGADYAESGSKSLKINETSFGVDDTVTDLFSD 75
Db 950 ALLVAEQENDLTKKAHANAQERNEELS--KEVEDADGKIK-----QLSDTVQRL--- 997

QY 76 KRTTPEKTIQNLAKGPREQELKAVTENTESKQITSGSQLSQSKESLS--LNKTVPTSTGNW 134
Db 998 BETIQEREALLAABRQEKKEEASAVIABSQARNE--AFASKLEDAEKQIDLLQETV---ORF 1053

QY 135 EICDPITKGNLTAVGLSKSGVEKLSQTDHLVPSQA--ADGTOLIOVASPFTPDKKTATAE 193
Db 1054 E--EAITKQLQSV-----TIEKQHEETVQVLAQAQAKIDELLREA---GDTDEKSTQLET 1104

QY 194 YTSRAGENGETSOLDV-----DGKEIINGEVFNSYLLKKKVTIPTGYKHIGQDAF 243
Db 1105 TIQRLSES--LITENDALLTTERQETATKLSLSEAQYKNBELLKKI-----EDA-- 1151
```



```

Db      1336  TEKENLVIANSKLNREELEHQNKLNELOVALAAVAEKEGSEEHSLKTLTDCMIQRKE 1395
QY      706   TTQOTFRDMTTAGKALLSKSNLROGEKQKFLQRAQFFLGRVDLDKAIKAEKALVTKKAT 765
Db      1396  ELESQVSSVTEHEELSKSYNTLSEKQ--LLNEKYESAKEGELGEAIAKLEBQNMVDKSE 1453
QY      766   KNGQL--LERSINKAVLAYNNSAIIKKNVRLKLEKLDLLTGLVEGKGPLAQATMVGUYL 823
Db      1454  KELHLSKLERQITLSBELTKMEB-IOTMQVETTEKDEALTTKME--HANLVH---- 1502
QY      824   LKTPLPLPEYYIGLVNFDKSKLIYALDMSDTIGEGOKDAYGNGPILNVDED-NEGYSHAL 882
Db      1503  -----EKOALEQQLEVRKELNDAYH-- 1523
QY      883   AVATLADVEGLDIKILNSKLSQLSIRQVPTAAVHR--AGIFQAIQNAAEAE 934
Db      1524  ---TIANQE-----EQSVREIKMDAYKKTSEDQLEAEQQHVABE 1561

```

## RESULT 22

```

US-10-282-122A-70177
Sequence 70177, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Lianguo
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zvekind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

```

```

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 70177
LENGTH: 3533
TYPE: PRF
ORGANISM: Staphylococcus aureus
US-10-282-122A-70177

```

```

Query Match      4.1%; Score 207; DB 15; Length 3533;
Best Local Similarity 19.2%; Pred. No. 0.004;
Matches 231; Conservative 169; Mismatches 424; Indels 380; Gaps 52;

```

```

QY      18   VTHNQEVPSLVKEPILKQTQASSISGADYAESGSKSL-----KINETSG----- 63
Db      1515  VONAESIAGLNNPTINK-----GNVSSATQAVISSKNALDQGVRLAQDQKQAGNSLNHL 1570
QY      64   --PVDDTVTDLFDKRTTPEKIKDNLAKGPREQ-LKAVTENTSESEKQITSGSQL----- 115
Db      1571  QLTFAQQALENQINNATTRDKVAEIIAQAAALNEAMKALKESIKDQPTQESKFINED 1630
QY      116  BOSKESISL-----LNKTVPTSTNWEICDFTTKGNTLVGLSKSGVEKLSQTDHLVLP 166
Db      1631  QAQDAVTAQVQAHAKDLINKTTPDLAKSIIDQATQAVTDAKNLHLGDKLQAQ-----D 1684
QY      167  SOAADGTQLIOVAFATPPDKKTAIAEYTSAGENGESQLDVGKEIINEGEVFNISYLL 226
Db      1685  KORA--TETLNNLSNLTTPQOQ-ALENQINNAATRGEVAQKLTEAQAALNQAMEALRSIQ 1741
QY      227  KKVITPTGYKHIGQD-----AFVQNK--IAEVNLP-----ESETISDYAFAPHL 269
Db      1742  DQOQTEAGSKFINEDKPKQDAYQAQVQNAKDLINQTNPTLDKAQVEQLTOAVNQAKDNL 1801
QY      270  ALKQIDLPDNLKAI GELAFDNIQTGKLSLPRQLMLAERAFKSNHINTIFRGNLSKVI 329
Db      1802  HGDQKLADDDKQHAVTDL-----NQLNGLNPNQOQAL-----ESQINNAATRGEVAQKL 1849
QY      330  GEASFQNDLSQL--MLPDGLEKTESAFTG--NPGDDHYNRVVLVTKSGKNSFGSLATE 385
Db      1850  ABAKALQAMQALRNSIQDQOQTESGSKFINEDKPKDAYQAQV-----QNAKDLINQ 1902
QY      386  NTVNPDKSLMOESPEIDYTKWLBEDFTYQKNSVTGFSNKGLOKVKRKN----- 434
Db      1903  TGNPTLDSQVEQ-----LTQAVTTAKDNL-----HGDQKLARDQQAQVTTVNALP 1948
QY      435  NLEIPKQ-----HNGVTITEIGDNAPRVNVDQNKTLRYDLEBEVLKPSIRKIGAPAF 487
Db      1949  NLNHAQQQALQATAINAAPTRTVEAHV-----QTATELDHAMETLK--NKVDQVNTDKA 2000
QY      488  QSNRLKFEASDDLEIEKEGAF-----QATSDKQNAVQKAIADA-----EN 2150
Db      2001  QPNYTT---EASTDKKEAVDQALQAAESITDPTGNSNANKDAVDQVLTCLKSEKELNGNE 2057
QY      515  ETLEKDKLVTIGDAAAPHINHIYAILVPESVOETGRSAFRONGANNILFMSKVKTLCGM 574
Db      2058  RVAEAKTQAKQTIQDLTHLN-----ADQIATA--KQN-----IDQATKLOPIAEL 2100
QY      575  AFLSNRL-----EHLDSQKQLTEIPVQAFSD--NALKEVLLPASLKTIRREE 620
Db      2101  VDQATQLNQSMDQLQQA VNEHANVEQTVDYT---QADSDKQNAVQKAIADA-----EN 2150
QY      621  APKQHLKQLEVASALSHIAFNALDNDGDGDFQNKVVVVKTH--H-----NSYALADGEH 673
Db      2151  VLKQNAKQ-QVDQALQNI-LNAKQALNGDERV---ALAKTNGKHDIQDLNALNNAQODG 2205
QY      674  F-----IVDP-----DKLSSTIVDL----- 688
Db      2206  PKGRIDQSDNLNQIQIIVDEAKALNRAMDQSIBITNEGRTKGSTNYNADTVQKVQYD 2265
QY      689  EKILKLEGLDYSTLRQTTQ---PRDMTTAGKALLSKSNLRQGEKQKFLQ----- 737
Db      2266  ETVDKAKQALDKSTGQNLTAQVILKNDVATAAKKALNGEBRLNRRKAEALQRLDQLTHL 2325
QY      738  ---EAQFPLGRVDLDKAIKAEKALVTKKATKNGQLL---ERSINKAVLA----- 781
Db      2326  NNAQROLAIQQINNAETLNKASRAI--NRATKLDNAMSGVQQYIDRQHLGVISSTNYINA 2383
QY      782  -----YNN-----SAIKKANVKELEKELDLTGLVEGKGPLAQATMVGQ 820
Db      2384  DDNLKANYDNAIANAAHLEDKVCQNAKAEAEQKQNIIDAQNALNGDQNLAK----- 2439
QY      821  VYLLKTPPLPEYYIGLVNVPDKSGKLIYALDMSDTIGEGOKDAYGNGPILNVDE----- 874
Db      2440  -----DKANAFVNSL---NGLNQOQOQDLAHKAINNADTVSDVTD 2475

```

```
QY 875 -----DNEGYYHALAVATLADYEGLD-----IKTILASK 902
Db 2476 IVNNQIDLDNAMEDTLKHLVDNEIPNA---EQVYVYNQADNNAKTNFDDAKRLANTLLNSD 2532
QY 903 LSQLTSIR---QVPTAAYHRAGIFQAIQNAABAEQQLPKPGTHSEKSSSESANSGKRG 959
Db 2533 NTNVDINGAIOAVNDIHNGLNGDORLODAKDKATQISNOALANKLKEIASNATDQDKL 2592
QY 960 LQSN 963
Db 2593 IAKN 2596

RESULT 23
US-10-470-048B-81
; Sequence 81, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; FILE REFERENCE: ANTIGENS TO A SPECIFIC PATHOGEN
; CURRENT APPLICATION NUMBER: US/10/470,048B
; CURRENT FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Scaphylococcus aureus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (1222)..(1225)
; OTHER INFORMATION: X = anything
US-10-470-048B-81

Query Match 4.1%; Score 206.5; DB 17; Length 1992;
Best Local Similarity 19.0%; Pred. No. 0.0019;
Matches 220; Conservative 157; Mismatches 420; Indels 361; Gaps 48;

QY 46 DYAESGSKKLINETSQVDTVDLFDKRTTPEKIKDNIAKGRBOELKAVNTES 105
Db 19 DHAKOTVSQLAHNNAQKHEMTLI-----DSETTRVAKQDLTEAQAQLDQALDQQA 74
QY 106 EKOITSGSQLEQSKESLSLNTKVPSTSNWEICDFITKGNLTGLSKSGVEK---LSOTDH 162
Db 75 DKDATRASSAYVNAEPNKKQSDVEAVQNAE-----SIIAGLNPTINKGNVSSATQA 126
QY 163 LVLPSQAADGTQLI-----QVASFA-----FTPDKKTATAEYTSRAGENGESQLDVGK 212
Db 127 VISSKNALDGVRLAQDKQTAGNSLNHLDLQTPAQQAQALLENQINNATTRGEVAQKLEAQ 186
QY 213 EIINGEVNSYLLKKVTPTGYKHIGQD-----AFVDNKN---IAEVNLP----- 255
Db 187 ALNQAEMALRNSIQDQOQTEAGSKFNEKPKDAYQAAVQNAKDLINQTNPTLDKAGV 246
QY 256 ESETISDYAFALHAKLDLPDLNKAIGELAFFNQITGKLSLPLRLMLRAERAFKSNH 315
Db 247 EQLTQAVNAQKNLHGDKQKLADKQHAVTDL-----NQLNGLNPNQRAQAL-----ESQ 294
QY 316 IKTIFRGNLSKIVIGEASPDNDLSQ--MLPDGLEKIESEAFGTG--NFGDDHYNNRVVL 371
Db 295 INNAATRGEVAQKLAELAKALDQAMQALRNSIQDQOQTESGSKFNEKPKDAYQAAV-- 352
QY 372 WTKSGNPNGLTENTYNNPDKSLMOBSEIDYTKWLEEDFTYQKNSVTGFSNKGQKVK 431
Db 353 -----QNAKDLINQTNPTLDKQSVQEQ-----LTQAVTTAKDNL-----HGPQKLA 393
QY 432 RNK-----NLEIPKQ-----HNGVTITEIGNAFNRVDFQNKLRKLDLEV 473
Db 394 RDOQQAQVTTVNALPNLHNAQQAQALTDAINAAPTRTEVAQHV-----QTATELDAME 447
QY 474 KLPSTIRKIGAFQAFQSNLNKSPFASDDLEIKEGAP----- 509
```

```
Db 448 K--NKVDQVNTDKAQPNYT---EASTDKKEAVDQALQAESITDPTNGSNANKDAVDQVL 502
QY 510 -----MNNRIETLEKDKLVITIGDAAFHINHIYAIVLPESVQEIGRSFRONGANN 560
Db 503 TKLOKELNGNERVAEAKTOAKOTIDOLTHLN-----ADQIATA--KQK---- 546
QY 561 LIFMGSKVKTIGEMAFLSNRL-----EHLDLSEQKQLTEIPVQAFPSD--NALK 606
Db 547 -IDQATKLOPIAELVDQATQLNQSMDLOQAVNEHANVEQTVDYT-----QADSDKQAYK 601
QY 607 EVLLPASLKTIRREAFKKNHLKQLEVASALSHIAFNALDDNDGDQFQNKVVVKT--H- 663
Db 602 QAIADA-----ENVLKQANKQ--QVDQALQNI--LNAKQALNGDERV---ALAKTNGKHD 650
QY 664 ---NSYALADGEHF-----IVDP-----DKLSSTIVDL----- 688
Db 651 IDQLNALNNAQODGPKGRIDQSDNLNQIQIVDEAKALNRAMDQLQSQETIDNEGRTKGST 710
QY 689 -----EKILKLIEGLDYSTLRTTOTQ---FRDMMTAGKALLSKSNLRQGE 731
Db 711 NYVNADTVQKVYDETVDKAKQALDKSTGQNLTAQVILKNDAVTAAKKALNGEERLNNR 770
QY 732 KQKFLQ-----EAQFFLGRVLDLKAIAKAEKALVTKKATNGQLL---ERSINK 777
Db 771 KAEALQRLDQLTHLNNAQRLAQIQQINNAETLNKASRAI--NRATKLDNAMGAVQOYIDE 828
QY 778 AVLA-----YNN-----SAIKKANVVKLEKELDLTLGLV 806
Db 829 QHLGVISSTNYNADNKLKANYDNAIAANAHELDKVGNAIAKAEAEQKQNIIDAQNAL 888
QY 807 EGKGPLAQATVQGVYLLKTPLEPYEYIYGLNVYFDKSKGLIYALDMSDTTIGEGOKDAYG 866
Db 889 NGDQNLNAK-----DKANAFVNSL---NGLNQOQODLAH 920
QY 867 NPILNVDE-----DNEGYYHALAVATLADYEGLD----- 894
Db 921 KAINNADTVSDVTDIVNNQIDLDNDAEMETLKHLDVNEIPNA---EQVYVYNQADNNAKTNF 977
QY 895 -----IKTILNSKLSQTSIR---QVPTAAYHRAGIFQAIQNAABAEQQLPKPGTHSE 945
Db 978 DDAKRLANTLLNSDNTVNDINGAIOAVNDIHNGLNGDORLODAKDKATQISNOALANKL 1037
QY 946 KSSSESANSGKDRGLQSN 963
Db 1038 KEIASNATDQDKLIAKN 1055

RESULT 24
US-09-815-242-5834
; Sequence 5834, Application US/09815242
; Patent No. US20020061589A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
```

```
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5834
; LENGTH: 2437
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-5834

Query Match      4.1%; Score 206.5; DB 9; Length 2437;
Best Local Similarity 19.0%; Pred. No. 0.0026;
Matches 220; Conservative 157; Mismatches 420; Indels 361; Gaps 48;

QY 46 DYASSGSKKLKINTSGVDDTVTDLFSDKRTTPEKIKONLAKGPRQELKAVNTES 105
DB 543 DHAKQTVSQLHLNNAQKHMEDTLI---DSETRTAVKQDLTEAQAALDQLMDALQQA 598
QY 106 EKQITSGGLEQSKESLSLNTKTPSTSNWEICDFITKGNLTGLSKSGVEK---LSQTDH 162
DB 599 DKDATRASAYVNAEPNKKQSYDEAVQNAE-----SITAGLNNPTINKGNVSSATQA 650
QY 163 LVLPSQAADGTQLI---QVASFA-----FTPDKKTAIAEYTSRAGENGEISQLDVDGK 212
DB 651 VISSKNALDGVRLAQDKQTAGNSLNHLDLTPAQQAQALENQINNATTRGEVAQKLETAQ 710
QY 213 EIINGEVFNISVLLKKVTIPTGYKHIGQD-----AFVDNKN---IAEVNLP----- 255
DB 711 ALNQAMEALRNSIQDQQQTEAGSKFINKEDPKQDAYQAQVQNAKDLINQTNPTLDKQAV 770
QY 256 BSLETISDYAPAHKALKQIDLDNLKAGELAFPNQITGKLSLPRQLMRLAERAFKSNH 315
DB 771 EQLTAQVNAKONLHGDKLADKKQHAVTDL-----NOLNGLNNPQOAL-----BSQ 818
QY 316 IKTIIFRGNLSKIVGEASFQDNDLSQL--MLPDGLEKIESEAFSG--NPGDDHYNRNVVL 371
DB 819 INNAATRGVAKLAERAKALDQAMQALRNSIQDQQTESGSKFINKEDPKDAYQAAV-- 876
QY 372 WTKSGNPSGLATENTYNNPDKLSNQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVK 431
DB 877 -----QNAKDLINQTNPTLDKRSQVEQ-----LTQAVTTAKDNL-----HGDQKLA 917
QY 432 RNK-----NLEIPKQ-----HNGVTITEIGDAPFRNVDPNKTLRKDYDLEEV 473
DB 918 RQQAQAVTTVNALPNLHAQQAALTDAINAAPTREVAQHV-----QTATELDHAMEYL 971
QY 474 KLPSTIRKIGAPAFOSNNLKSPEASDDLEEIKEGAF-----ADQIATA--KQN----- 509
DB 972 K--NKVDQVNTDKAPNYT---EASTDKKEAVDQALQAESITDPTNGSNANKDAVDQVL 1026
QY 510 -----MNNRIETLEKOKLVITGDAAFHNNHYAIVLPESVOEIGRSFRONGANN 560
DB 1027 TKLQEKENELNGERNVAEAKTOAKQOTIDQLTLN-----ADQIATA--KQN----- 1070
QY 561 LIFMGSKVKTGEMAFLSNRL-----EHLDLSEKQKLTETPVOAFSD--NALK 606
DB 1071 -IDQATKQPIAELVDQATQLNQSDMDQLQAAVNEHANVEQTVDYT-----QADSDKONAYK 1125
QY 607 EYLLPASLKTIREBAFKKHLKQLEVASALSHAFNALDDNDGDQFDNKVVVKTTH--H- 663
DB 1126 QAIADA-----ENVLKQNAKQ-QVDQALQNI--LNKAQALNGDERV---ALAKTNGKHD 1174
QY 664 -----NSYALADGHP-----IVDP-----DKLSSTIVDL----- 688
DB 1175 IDQLNALANAQDGFGRIDQSDNLDNLQIQIIVDEAKALNRAMDQLSQEITDNEGRKTGST 1234
QY 689 -----EKILKLIBGLDYSLRQTTQTOQ---FRDMTTAGKALLSKSNLRQGE 731
```

```
DB 1235 NYVNADTVQKVQVYDVTVDKAKQALDKSTGQNLTKAQVIKLNDVATAAKKALNGERLNRR 1294
QY 732 KQKFLQ-----EAQFFLGRVDDDKAIAKAERKALVTTKATQNGQLL---ERSINK 777
DB 1295 KAEALQRLDQLTHLNAQRQLAIQIINNAETLNKASRAI--NRATKLDNAMGAQQYIDE 1352
QY 778 AVLA-----YNN-----SAIKKANVRKLEKELDLTLTGLV 806
DB 1353 QHLGVISSTNYINADNKLKANYDNANAAHELDKVOGNATAKAEAEQLKQNIIDAQNAL 1412
QY 807 EGKGPLAATWVGYYLLKTPPLPYEYIGLVVFDKSGKLIYALDMSDTTIGEGOKDAYG 866
DB 1413 NGDONLANAK-----DKANAFVNSU---NGLNQOQOOLAH 1444
QY 867 NPILNVDE-----DNEGYPHALAVATLADYEGLD----- 894
DB 1445 KAINNADTVSDTVINNQIDLNDAMEFTLKHLDVNEIPNA---EQTVNYQNADDNAKTNF 1501
QY 895 -----IKTILNSKLSQTSIR---QVPTAYHRAGIFQAIONAAAEABQLLPKPGTHSE 945
DB 1502 DDAKRLANTLLNSDNTVNDINGAIOAVNDAIHNLGDRLOQDAKDKAIOGINQALANKL 1561
QY 946 KSSSESSEANSKDRGLQSN 963
DB 1562 KEIEASNATDQDKLIAKN 1579
```

## RESULT 25

```
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zvakind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12996
; LENGTH: 6281
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996
```

```
Query Match      4.1%; Score 206.5; DB 9; Length 6281;
Best Local Similarity 19.0%; Pred. No. 0.0097;
Matches 220; Conservative 157; Mismatches 420; Indels 361; Gaps 48;

QY 46 DYASSGSKKLKINETSQVDDTVTDLFSDKRTTPEKIKONLAKGPRQELKAVNTES 105
```

Db 4385 DHAKQTVSOLAHNNNAQKHMEDTLI-----DSETRTAVKQDLTEAQALDQALMDALQOOSTA 4440  
Qy 106 EKQITSGSQLESKESLSLNKTVPTSTNWEICDFITKGNLTVLGSKSGVEK---LSOTDH 162  
Db 4441 DKDATRASAYVNAEPNKKQSYDEAVQNAE-----SIIAGLNNPTINKGNVSSATQA 4492  
Qy 163 LVLPQQAADGTOLI-----QVASFA-----FTPDKKTATIAEYTSRAGENGESQLDQVGGK 212  
Db 4493 VISSKNALDQVERLAQDKQTAGNSLNHLDLTPAQQAQALNNATTRGEVAQKLTQEAQ 4552  
Qy 213 EIINEGEVNSVLLKKVTIPTGYKHIGQD-----AFVDNKN--IAEVNLP-----255  
Db 4553 ALNQAMEALRNSIQDQOQTEAGSKFINEDKPKQDAYQAVQNAKOLINQTNPTLDKAQV 4612  
Qy 256 ESLETTISDYAFALHAKLQIDLPNLKAIGELAFDQITGKLSLPQLMRLAERAPKSNH 315  
Db 4613 EQLTQAVNQAKDNLHGDKLADDKQHAVTDL-----NQLNGLNPNQOQAL-----ESQ 4660  
Qy 316 IKTIIEFRGNSLVKIGEASFQDNDLSQL--MLPDGLEKIESEAFGTG--NFGDDHNNRVVL 371  
Db 4661 INNAATRGEVAQKLAEAQALDQAMQALRNSIQDQOQTESSGSKFINEDKPKQDAYQAAV--4718  
Qy 372 WTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGKQVK 431  
Db 4719 -----QNAKDLINQTNPTLDKSKVBEQ-----LQAVTTAKDNL-----HGDQKLA 4759  
Qy 432 RNK-----NLEIPKQ-----HNGVTITEIGDPAFNVDFQNKTLRKDYDLEEV 473  
Db 4760 RQOQAVTTVNALPNLHQAQQALTDAINAAPTRTEVAQHV-----QTATELDHAMEYTL 4813  
Qy 474 KLPSTIRKIGAFQFQSNLKSFEASDDLEEIKEGAF-----509  
Db 4814 K--NKVDQVNTDKAQPNYT---EASTDKKEAVDQALQAAESITDPTNGSNANKDAVDQVL 4868  
Qy 510 -----MNRITELBLKDLVTIGDAAFHINHIYALVLPESVQEIQRSPONGANN 560  
Db 4869 TKLQEKENELNGERNVAEAKTQAKOTIDQLTHLN-----ADQIATA--KQN-----4912  
Qy 561 LIFMGSKVITGEMAFPLSNRL-----EHLDLSEQKOLTEIPVQAFSD--NALK 606  
Db 4913 -IDQATKLOPIAELVDQATQLNQSMDQLOQAVNEHANVEQTVDYT-----QADSDKQNAKY 4967  
Qy 607 EYLLPASLKTIREEAFKGNHKLQEVASALSHIAFNALDDNDGDEQFNDKNVVKTH--H- 663  
Db 4968 QAIADA-----ENVLKQNAKQ-QVDQALQNI-LNAQALNGDERV---ALAKTNGKD 5016  
Qy 664 ----NSYALADGEHF-----IVDP-----DKLSSTIVDL-----688  
Db 5017 IDQLNALNNAQDQGFGRIDQSNDLNQIOQIVDEAKALNRAMDQLSQEITDNEGRTKGST 5076  
Qy 689 -----EKILKLEGLDYSTLRQTTQTO---PRMTTAKGALLSKSNLRQGE 731  
Db 5077 NYNADTQVKQVYDVTVDKAKQALDKSTQGNITAKQVINKINDAVTAAKALNGEERLNR 5136  
Qy 732 KQKFLQ-----EAQFELGRVLDLKAIAKAEKALVTKKATKNGQLL---ERSINK 777  
Db 5137 KAEALQRLDQLTHLNAQHQALQIQQINNAETLNKASRAI--NRATKLDNAMAQVOYIDE 5194  
Qy 778 AVLA-----YNN-----SAIKKANVYKLEKELDLTLGIV 806  
Db 5195 QHLGVISSTNYINADDNLKANYDANIAANAHELDKVQGNIAKAEAEQLKQNIIDAQNAL 5254  
Qy 807 ECKGPLAQATWQGVYLLKTPPLPEYYIGLVNYPDKSGKLIYALDMSITIGEGQDAYG 866  
Db 5255 NGDQNLANAK-----DKANAFVNSL---NGLNQOQOQDLAH 5286  
Qy 867 NPILANDE-----DNEGTHALAVATLADYEGLD-----894  
Db 5287 KAINNADTVSDVTDIVNNQIIDLNDAMETLKLHVDNEIPNA---EQVTYQNAADDNAKTNF 5343  
Qy 895 -----IKTILSKLSQLTSIR---QVPTAAVHRAGIFQAIQNAAAEAQLPKPETHSE 945  
Db 5344 DDAKRLANTLLNSDNTNVNDINGAIQAVNDAIHNLNGDRLQADKAKAIQSIQNALANKL 5403

Qy 946 KSSSESANSKDRGLQSN 963  
Db 5404 KEIEASNATDQDKLIAKN 5421  
RESULT 26  
US-10-470-048B-440  
; Sequence 440, Application US/10470048B  
; Publication No. US20050037444A1  
; GENERAL INFORMATION:  
; APPLICANT: MEINKE ET AL.  
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF  
; FILE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN  
; FILE REFERENCE: SONN:035US  
; CURRENT APPLICATION NUMBER: US/10/470, 048B  
; CURRENT FILING DATE: 2003-07-25  
; NUMBER OF SEQ ID NOS: 603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 440  
; LENGTH: 10498  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (9728)..(9731)  
; OTHER INFORMATION: X = anything  
US-10-470-048B-440

Query Match 4.1%; Score 206.5; DB 17; Length 10498;  
Best Local Similarity 19.0%; Pred. No. 0.02;  
Matches 220; Conservative 157; Mismatches 420; Indels 361; Gaps 48;  
Qy 46 DYAESGSGSKLKNETSGPVDVTVDLPSDKKTTTPEKIKONLAKGPREQELKAVNTES 105  
Db 8525 DHAKQTVSOLAHNNNAQKHMEDTLI-----DSETRTAVKQDLTEAQALDQALMDALQOOSTA 8580  
Qy 106 EKQITSGSQLESKESLSLNKTVPTSTNWEICDFITKGNLTVLGSKSGVEK---LSOTDH 162  
Db 8581 DKDATRASAYVNAEPNKKQSYDEAVQNAE-----SIIAGLNNPTINKGNVSSATQA 8632  
Qy 163 LVLPQQAADGTOLI-----QVASFA-----FTPDKKTATIAEYTSRAGENGESQLDQVGGK 212  
Db 8633 VISSKNALDQVERLAQDKQTAGNSLNHLDLTPAQQAQALNNATTRGEVAQKLTQEAQ 8692  
Qy 213 EIINEGEVNSVLLKKVTIPTGYKHIGQD-----AFVDNKN--IAEVNLP-----255  
Db 8693 ALNQAMEALRNSIQDQOQTEAGSKFINEDKPKQDAYQAVQNAKOLINQTNPTLDKAQV 8752  
Qy 256 ESLETTISDYAFALHAKLQIDLPNLKAIGELAFDQITGKLSLPQLMRLAERAPKSNH 315  
Db 8753 EQLTQAVNQAKDNLHGDKLADDKQHAVTDL-----NQLNGLNPNQOQAL-----ESQ 8800  
Qy 316 IKTIIEFRGNSLVKIGEASFQDNDLSQL--MLPDGLEKIESEAFGTG--NFGDDHNNRVVL 371  
Db 8801 INNAATRGEVAQKLAEAQALDQAMQALRNSIQDQOQTESSGSKFINEDKPKQDAYQAAV--8858  
Qy 372 WTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGKQVK 431  
Db 8859 -----QNAKDLINQTNPTLDKSKVBEQ-----LQAVTTAKDNL-----HGDQKLA 8899  
Qy 432 RNK-----NLEIPKQ-----HNGVTITEIGDPAFNVDFQNKTLRKDYDLEEV 473  
Db 8900 RQOQAVTTVNALPNLHQAQQALTDAINAAPTRTEVAQHV-----QTATELDHAMEYTL 8953  
Qy 474 KLPSTIRKIGAFQFQSNLKSFEASDDLEEIKEGAF-----509  
Db 8954 K--NKVDQVNTDKAQPNYT---EASTDKKEAVDQALQAAESITDPTNGSNANKDAVDQVL 9008  
Qy 510 -----MNRITELBLKDLVTIGDAAFHINHIYALVLPESVQEIQRSPONGANN 560  
Db 9009 TKLQEKENELNGERNVAEAKTQAKOTIDQLTHLN-----ADQIATA--KQN-----9052



```
Qy 561 LIFMGSKVKTGEMAFSLNRL-----EHLDSLSEKQLTEIPVQAFSD--NALK 606
Db 9053 -IDQATKLOPIAELVDQTLNQSMDQLQAVNEHANVEQTVDYT-----QADSDKQNAIK 9107
Qy 607 EYLLPASLKTREEAFKKNHLKQLEVASALSHIAFNALDNDGDGRQFNKVVVVTH--H- 663
Db 9108 QAIADA-----ENVLKQNAKQ-QVDQALQNI-LNAQALNGDERV---ALAKINGKHD 9156
Qy 664 ----NSYALAGEHF-----IVDP-----DKLSSTIVDL----- 688
Db 9157 IDQLNALNNAQQGFKGRIDOSNDLNLQIQIIVDEAKALNRAMDQLSQBITDNEGRTKGST 9216
Qy 689 -----EKILKLTLEGLDYSTLRQTTQTQ-----FRDMTTAGKALLSKSNLRQGE 731
Db 9217 NYVNADTVQVYDETVDAKQALDKSGQNIJAKQVTKLNDVAVTAAKKALNGBERLNR 9276
Qy 732 KOKFIQ-----BAQFPLGRVDLDKAIKAERKALVTKKATKNGQLL---ERSINK 777
Db 9277 KAEALQRLDQLTHLNAQRQLAIQQINNAETLNKASRAI--NRATKLDNMGAVQQYIDE 9334
Qy 778 AVLA-----YNN-----SAIKKANVKKLEKELDLGLTGLV 806
Db 9335 QHLGVISSTYINADNLKANYDNAIAAAHELDKVGQNAIAKABAEOLKQNIIDAQNAL 9394
Qy 807 EKGKPLAQTWQGVYLLKTPLPPEYYIGLVNVPDKSGKLIYALDMSDTIGEGQKDAYG 866
Db 9395 NGDQNLNAK-----DKANAFVNSL---NGLNQOQODLAH 9426
Qy 867 NPILNVDE-----DNEGYPHALAVATLADYEGLD----- 894
Db 9427 KAINNADTVSDVTDIVNNQIDLDAMWETLKLHLDVNEIPNA---EQTVNYQNADNAKTNF 9483
Qy 895 -----IKTILNKSQLTSTR---QVPTAAYHRAGIFQATQNAABAEQLLPKPGTHSE 945
Db 9484 DKAELANTLLNSDNTYNDINGAIAQAVNDATHNLNGDRLQDAKDKAIOQSIQNALANKL 9543
Qy 946 KSSSESANSKDRGILQSN 963
Db 9544 KEIASNATDQDKLIAXN 9561
```

## RESULT 27

```
US-10-282-122A-70580
; Sequence 70580, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Hatzelbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
```

```
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70580
; LENGTH: 6641
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6636)
; OTHER INFORMATION: X-any amino acid
US-10-282-122A-70580

Query Match 4.0%; Score 205; DB 15; Length 6641;
Best Local Similarity 19.1%; Pred. No. 0.013;
Matches 201; Conservative 178; Mismatches 447; Indels 224; Gaps 45;

Qy 47 YAESGSKLKINTSGV--DDTVTLFSDKRTTPEKIKDNLAKGPREBELKAVTENTE 104
Db 4050 YNQAVNAKAKNIINDQPTPVMANDEIQSVLNEVKQT---KDNL-----HGQDKLANDKTD 4100
Qy 105 SEKOITSGSQLESKESLSLNKTVSTSNMEICDFITKGNLTGLSGVGEKLSQTDHLV 164
Db 4101 AQTALNALYNLQNAQRGNLETKVQNSNSRSPVQKVQLANQLNDAMKLDLALTGNDAIK 4160
Qy 165 LPSQAADQTLQIVASFAFTPKKTAIAYTSRAGENGESQL-----DVGKEII 215
Db 4161 QTSYVINEDTSQVNFDEYTDGKNIVABQTNPNWSPNTINTIADKITEAKNDLHGQVKL 4220
Qy 216 NEGEVFNISLLKVT-IPTGYKHIGQDAFVGNKNIAEVN-----LPESLETISQYAF 266
Db 4221 KQAQQQSINTINQMTGLNQAQKEQLNQBIOQTOTRSEVHVQVINKAQAALNDSMTLRQSIT 4280
Qy 267 AHLAKQIDLPDLNKAIGELAPPN-----OITGKLSLPQLMLAEAFKSNHIKTIE 320
Db 4281 DEHEVKQTSNYIN-ETVGNQFAYNNADVDRVKQIINQTSNPTMNPLEVERA-----TSNVK 4334
Qy 321 PRGNLSKIVGEASPDNDLSQLMLP---DGLEKIESEAFGTG-----NPGDDHNNRV 369
Db 4335 ISKDALH--GERELNDKNKSKTFAVNHLDNLNQAQKEALTHEIEQATIVSQVNNIYNKAK 4392
Qy 370 VLMTKSGK-----NPSGLATENTYVNPDP---KSLMQE-----SPEIDYTKM 407
Db 4393 ALNNDMKKLKDIVAQDNVRSNNYINEDSTPQNNYNDNTINHAQSIIIDQVANPTMSHDE- 4451
Qy 408 LEEDPTQKNSVTGFSNK-GLQVKRNKNLEI PQHNGVTTITEIGDNAFRNVDPQNKTLR 466
Db 4452 IENAINNIKHAINALDGEHLQQAENANL-----LINSNLNLAPODAINRLVN 4502
Qy 467 KYDLEEVKLPSTIRKIGAFQSNLKSFEASDDELEEKEGAFMNNRIETLELKDCLVTI 526
Db 4503 EAQTRF-----KV-----AEQLQSAQALNDAMKHLRNSIQNQ--SSVRQESKYNA 4546
Qy 527 GDA-AFHINHIYAI-----VLPESVQEIGRSAFRQ-----NGANNLIPMGSKVKTGEMA 575
Db 4547 SDAKKEQYNH--AVREVENIINEQHPTLDKEIKQLTGVNQANNDL-----4591
Qy 576 FLNRLHLDLSEKQLTEIP---VQAFSDNALKE-----VLLPAS 613
Db 4592 ---NGVELLDADKQNAHQSIPTLMHLNQAQOONAKNEKINNNAVTRTEVAATIGQAKLLDHA 4648
Qy 614 LKTIREAFKKNHLKQLEVASALSHIAFNAL--DNDGDGEQFDN-----KVVVKTHNSY 666
Db 4649 MENLEESIKDKEQVKQ-----SSNYINSDSVQETVDNAVDHVTYILNQTVNPTL 4698
```



[illegible]

```
Db 1405 QDILNAQTNDVTOIKDQAVADTQGITADTTIKDVAK-----DELATKAN----- 1449
Qy 175 LIQVASFAPTPDKKTAIAEYTSRAGENGESQLDVGKEIINEGEVFNSEYLLKKVTIPTG 234
Db 1450 -----EQKALIAQTADATTEKEQANQOVDAQ-----LTQG 1480
Qy 235 YKHIGQDAFVNDKNIAEVLNPESELETISDYAFALHAKQIDLPDLNKLKAIGELAFFNQIT 294
Db 1481 NQNIENASQIDVNTAKDNAIQAIPIQ-----ASTDVKTNARA-ELLTEMQNKIT 1530
Qy 295 GKLSLPROLMRLAERAFKSHIKTIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESE 354
Db 1531 EIL-----NNNETTNEEGNDIGPV-RAAYE-----EGLANNIAA 1564
Qy 355 AFTGNPGDDHNNRVVLWTKGKNSGLATENTYVNPDKSLWQESPEIDYTKWLEBDFTY 414
Db 1565 TTIGD-----VTTAKDTAVQKVQQLHANPVKKPACK-----KELDQAAD 1604
Qy 415 QKNSVTGFGKGLQKVRNKNLEIPKOHNGVTITEIGDNAPRVDPQNKTKLYDLSEYVK 474
Db 1605 KKTQIBQTPNASQOEINDAKQ-EVDTELN-----QAKTNVD-QSSTNEYVD----- 1648
Qy 475 LPSTIRKIGAFAPQSNLKSFEASDDLEIKGAFMNNRIETLE--LKDKLVTIGDAAPH 532
Db 1649 -----NAVKEGKAKINAVKTFSEYKDKALAKIEDA----- 1678
Qy 533 INHIAVLPEVQETGRSAFRONGANNLIFMGSKVKTILGEMAFLSNRLEHL-----DL 586
Db 1679 -----YNAKVNEADN-----SNASTSSEIAEAKQKLAELKQADQNV 1715
Qy 587 SEQKQTEIPVAFSD-NALKEVLLPASLK-----TIREAPFK 624
Db 1716 NQATSKDDIEVQIHNDLNDINDYTIPTGKESATTDLYAYADQKKNISADTNATQDEQ 1775
Qy 625 NHLKQLE--VASALSHIAFNALDNDGDE-----QPDNKKVVVTHHNSVALA- 669
Db 1776 QAIKQVDQNVQTALENIN-NGVDNGVDVDDALTOGKAAIDTIQVDATVPKQANQAIKAE 1834
Qy 670 DGHFTVDPDKLSS--TIVDLKILKILGLDYSLRTQTTOTQFRDMMTAGKALLSKSL 727
Db 1835 DYKESIDHSDQLTAEKTEALAMIKQITQAKQGYTDATTTAEVEKAKAGLEAFDNIQI 1894
Qy 728 ROEKQKQFLOEQOFFLGR-----VLDKKAIAKAEKALVT-----KKATKNGQLLERSIN 776
Db 1895 DSTEKQKAEIELETALDQIEAGVNVVDADATTEBEKAFTNALEDILSKATED--ISDQTN 1952
Qy 777 KAVLAYNNSAIAKANKVRL-----EKELDLTLGLVEGKPLAQATWVG 820
Db 1953 AETATVKNSSALEOLKAQRINPVVKNKNALEAIREVNVNKQIEII-----KNADADASAK 2005
Qy 821 VYLLKTPPLPEYIIGLVVYFDK-SGKL-----IYAL-----DMS 854
Db 2006 --IART-----DLGRYDFRPAKLQKTQNTVEAELQNVTPAIEAIVPQNDPDAN 2054
Qy 855 DT-IGEGQKDAYGNPILANVDENEGHEHALAVLATLADYEGLDIKTILNSKLSOLTSIRQVP 913
Db 2055 DTNNGTDNDNATANSANATPENTGQPNVSEIT--DNKRADASPTTPNNSDAATGETT 2112
Qy 914 TAAHRAIGIFOAIONAAAEQLLP-----KPTHSEKSSSSSES 952
Db 2113 SATDDAKDPQANNSADASTNSPTMNDNVTSKPEVESTNNGTTDKPVTETDNATPAPS 2172
Qy 953 A---NSKDRGLQSNPKNTGRHSAILPRTGS 980
Db 2173 TTNNSTTTATNENAPTQ---STATAPTAS 2200
```

## RESULT 31

US-10-282-122A-43811

; Sequence 43811, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

```
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

```
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
```

```
; CURRENT APPLICATION NUMBER: US/10/282,122A
```

```
; PRIOR FILING DATE: 2003-02-20
```

```
; PRIOR FILING DATE: 2000-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
```

```
; PRIOR FILING DATE: 2000-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/206,848
```

```
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
```

```
; PRIOR FILING DATE: 2000-05-26
```

```
; PRIOR APPLICATION NUMBER: 60/230,335
```

```
; PRIOR FILING DATE: 2000-09-06
```

```
; PRIOR APPLICATION NUMBER: 60/230,347
```

```
; PRIOR FILING DATE: 2000-09-09
```

```
; PRIOR APPLICATION NUMBER: 60/242,578
```

```
; PRIOR FILING DATE: 2000-10-23
```

```
; PRIOR APPLICATION NUMBER: 60/253,625
```

```
; PRIOR FILING DATE: 2000-11-27
```

```
; PRIOR APPLICATION NUMBER: 60/257,931
```

```
; PRIOR FILING DATE: 2000-12-22
```

```
; PRIOR APPLICATION NUMBER: 60/267,636
```

```
; PRIOR FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
```

```
; PRIOR FILING DATE: 2001-02-16
```

```
; Remaining Prior Application data removed - See File Wrapper or PALM.
```

```
; NUMBER OF SEQ ID NOS: 78614
```

```
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 43811
```

```
; LENGTH: 6713
```

```
; TYPE: PRF
```

```
; ORGANISM: Staphylococcus aureus
```

```
; US-10-282-122A-43811
```

```
Query Match 4.0%; Score 203; DB 15; Length 6713;
```

```
Best Local Similarity 19.5%; Pred. No. 0.018; Mismatches 396; Indels 376; Gaps 64;
```

```
Matches 232; Conservative 183;
```

```
Qy 34 KQTQASSSTISGAD-----YAESGSKSLKINETSGP-VDDTVTDLFSKRTTPEKIKDN 86
```

```
Db 5050 QQTEAGSKFNEFKPKDAYQAAVQNAKDLINQTNPTLDKQVQLTQAVTT---AKDN 5106
```

```
Qy 87 L---AKGPREQLKAVTEN-----TESEKQITSSGSEQSKE 120
```

```
Db 5107 LHGDQKLARDQQAQAVTTVNALPNLHNAQQOQTLTDAINAAPTREVAQHVTATELDNAME 5166
```

```
Qy 121 SLSLNTKVPSTNWEICDFITKGNLTVLGSKSGVEKLQTOHHLVLPQQAADGTOLIOVAS 180
```

```
Db 5167 TLK-NKV-----DOVNTDKAQPNTTEASTDKKBAVD-----QALQAAQ 5203
```

```
Qy 181 FAFTPD-----KKTAAIABYTSRAGENGISQLDVGKEIINEGEVFNSEYLLKKVTIPTGY 235
```

```
Db 5204 SITDPTNGSNANKDAVEQALTKLOE--KVNEL--NGNERVAEAKTQAKTIDQLT----- 5254
```

```
Qy 236 KHIGQDAFVNDKN-----IAEV-----NLPSELETISDYAFALHAKQ-IDL--- 276
```

```
Db 5255 -HLNADQIATAKQNTDOATKLOPIAELVDQATQLNQSMQDLQQAQVNEHANVEQTIDYTOA 5313
```

```
Qy 277 -PDNLKAIGE-LAFFDN---QITGKLSLPRQLMRL--AERAF-----KSNHKTIE 320
```

```
Db 5314 DSDKQKAYQAIADAENVLKQNAKQOVDQALQNTILNAKQALNGDERVALAKTKHDDID 5373
```

```
Qy 321 FRGNSLKVIGEASF-----QDNDLSQLM-----LPDGLKIESEAPTGNPGDDHNN 367
```

[illegible]

RESULT 32

```

US/09-815-242-5703
/ Sequence 5703, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ TITLE OF INVENTION: Prokaryotes
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815.242

```

```
Db 985 INNAETLNKASRAINRATKLDNAMGAVQQYIDQHLGVISSTNYINA-----DDNLKANYD 1040
QY 495 ---FEASDDEETKEGAFMNNRIETLEKDKLVITIGDAAAFHINHIAIVLPESVQIGRS 551
Db 1041 NAIANAHELDKVOGNAIA--KAAEQKQNIIDAQNAL-----1077
QY 552 AFRONGANNLIPMGSKVKTGEMAFLSNRLEHLDLSEQKQLTEIPVOAFSDNALKEVLLP 611
Db 1078 ---NGDQLANAKDKAN-----AFVN-----SLNGLNQ-----1102
QY 612 ASLKTIRREAFKQKHLKQLEVASALSIAFNALDDNDGDEQF-----DNKVVVVKTHNSYA 667
Db 1103 ---QOQDLAHNAINNADTVSDVTIDVNNQIDLDNDAETLKLHLDVNDNIPNAEQTVNYQ 1156
QY 668 LADGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLRQTQTQFRDMTTAGKALLSKSNL 727
Db 1157 NAD-----DNKTNPDACKELA-----NTLLNSDNTVNDINGAIOAVKDAION 1200
QY 728 ROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTK-----KATKNGOLLERSINKA-V 779
Db 1201 LANGEQR--LQEAQ-----DKAIQNVNKLADKLKEIEASNATDQDKLIAK--NKABE 1248
QY 780 LANNNAIKKANVKRLEKELDLTLGLVEGKGLAQATMVQGVYLLKTPLPPEYVIGLVN 839
Db 1249 LA--NSIIN--NINKATSNDQVSQVQTAGNQAIEQVHANE-----IPKAKIDANK 1294
QY 840 YFDKSGK-LIYALDMSDTTIGEGOKDAYGNPILNVDED--NEGHALAVATLADYEG--- 892
Db 1295 DVDKQVOALIDEIGRPNLTDRKQKALDRINGILOQHNDINNALTKEAIEQAKERLAQ 1354
QY 893 --LDIKTIILNSKLSQTSIRQVPTAAHYHRAGIPQAIQNAAAEAQQLPKPGTHSEKSSS 950
Db 1355 ALQDQIKOLVKAKEDAKNKIKAL-----ANAKRDQINSNPDLTPEQKAKA 1398
QY 951 --ESANSKDRGLQS--NPKT-----NRG 969
Db 1399 LKEIDEAEKRALQNVENATQIDQNLNRG 1425
```

## RESULT 33

```
US-09-815-242-12611
; Sequence 12611, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 12611
; LENGTH: 3158
; TYPE: PRN
; ORGANISM: Staphylococcus aureus
US-09-815-242-12611
```

```
Query Match 4.0%; Score 202.5; DB 9; Length 3158;
Best Local Similarity 18.5%; Pred. No. 0.0065;
Matches 216; Conservative 174; Mismatches 410; Indels 367; Gaps 55;

QY 16 SVVTHNQEVPSLVKPEILKQTOASSISGADVAESS--CKSKL-----KINET 61
Db 1544 AAVQKAKDLINTSPTLDKQAEQVLTQGVNAKQNLHGDQKLDDKQHAVVDLNLQNSL 1603
QY 62 SQPVDDTVTDLPSDKRTTPEKIDNLAKGPR--EQLKAVTENTSESKQITSGSOLQSKE 120
Db 1604 NNPQRALESQINNAATRDE--VAQKLAEAQALDQAMQALRNSIQDOOQTESSEK 1657
QY 121 SLSLNTVPSTSNWIC-----DFITK--GNTVLGLSKGVEKLSQT-----DHL 163
Db 1658 ---INEDKPKDAYQAAVQKAKDLINQTNPT--LDKSOVEQLTQAVTTAKDNLHGDQKL 1712
QY 164 -----VLP-----SQAADGTQLIQVASFAPTPDKKTAIAEYTSRAGE----- 200
Db 1713 ARDQQAQVTTVNALPNLHAQQAALTDAINAA-----PTRTEVAQHVTATFELDHAMET 1766
QY 201 -NGEISOLD-----VDGKEINEGEVFNFSYLLKKVTIPTGYKHIQODAFVD-- 245
Db 1767 LKKVDQVNTDKAQPNYTEASTDKKEAVDQA-----LQAAESITDPTNGSNANKDAVEQAL 1822
QY 246 -----NKNIAEYNLPESLETISDYAFHL-----ALKQIDLPNLKAIGEL 286
Db 1823 TKLQEKVNLNGNERNVABAK--AAQKQITDQ--LAHLNADQIATAKQNIQATKQPIAEL 1879
QY 287 AFPDNQI-----TGKLSLPQMLRAERAFKSNHIK-- 317
Db 1880 VDAQTLNQSMQDLOQA VNEHANVEQTVDTQADSDKQNAKQAEAEVNLKQNSKQ 1939
QY 318 -----TIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKTESEFTNGPDHYN 366
Db 1940 VDAQALQNLINAKQALNGDERVALAKTNGK--HOIDQJ---NALNNAQDQFGKRIDQSHDL 1995
QY 367 NRVVLWTKSGK-----NPSGLATENTYVNPDKSLWQESPE-IDYTKW--- 407
Db 1996 NQIQIIVDEAKALNRAMDLSOEISGNEGRTKSTNYVNAQTQVKQYVDEAVDKAKALD 2055
QY 408 -----LEEDPTYQKNSVTG--FSNKGLOKQVR-----NKNLEIPK 440
Db 2056 KSTGQNLTAEQVILKNDVAVTAAKKALNGEERLNKRKSEALQRLDQLTHLNNARQLAIQ 2115
QY 441 CHNGVTITEIGD--NAFRNVDFQNTLRKY--DLEEVKLPSTIRTKTGAPAFQSNLKS--- 494
Db 2116 INNAETLNKASRAINRATKLDNAMGAVQQYIDQHLGVISSTNYINA-----DDNLKANYD 2171
QY 495 ---FEASDDEETKEGAFMNNRIETLEKDKLVITIGDAAAFHINHIAIVLPESVQIGRS 551
Db 2172 NAIANAHELDKVOGNAIA--KAAEQKQNIIDAQNAL-----2208
QY 552 AFRONGANNLIPMGSKVKTGEMAFLSNRLEHLDLSEQKQLTEIPVOAFSDNALKEVLLP 611
Db 2209 ---NGDQLANAKDKAN-----AFVN-----SLNGLNQ-----2233
QY 612 ASLKTIRREAFKQKHLKQLEVASALSIAFNALDDNDGDEQF-----DNKVVVVKTHNSYA 667
Db 2234 ---QOQDLAHNAINNADTVSDVTIDVNNQIDLDNDAETLKLHLDVNDNIPNAEQTVNYQ 2287
QY 668 LADGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLRQTQTQFRDMTTAGKALLSKSNL 727
Db 2288 NAD-----DNKTNPDACKELA-----NTLLNSDNTVNDINGAIOAVKDAION 2331
QY 728 ROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTK-----KATKNGOLLERSINKA-V 779
Db 2332 LANGEQR--LQEAQ-----DKAIQNVNKLADKLKEIEASNATDQDKLIAK--NKABE 2379
```



QY 780 LAYNSAUKANVKELEKELDLTGLVSGKGLAQATWQGVYLLKTPPLPEYVIGLVN 839  
Db 2380 LA--NSIIN--NINKATSNQDVSQVOTAGNQAEQVHANE--IPKAKIDANK 2425  
QY 840 YPDKSGK-LIYALDMSDTIGEGQKDAYGNPILNVDED--NEGYPHALAVATLADYEG-- 892  
Db 2426 DVDKQVALDEIGRPNLTDEKQALKDRINQILQOQHNDINNALTKEATEQAKERLAQ 2485  
QY 893 --LDITKLNSKLSQTSIRQVPTAAYHRAGIFQAIQNAABAEQOLLKPKGTHSEKSSSS 950  
Db 2486 ALQDQIKVLKAKEDAKNIKAL-----ANAKRDQINSNPOLTPBOKAKA 2529  
QY 951 --ESANSKDRGLQS--NPKT-----NRG 969  
Db 2530 LKEIDAEKRALQNVENAQTIDQLNRG 2556

## RESULT 34

US-10-994-726-322  
; Sequence 322, Application US/10994726  
; Publication No. US20050147999A1  
; GENERAL INFORMATION:  
; APPLICANT: NI et al.  
; TITLE OF INVENTION: Lyme Disease Vaccines  
; FILE REFERENCE: PB481DI  
; CURRENT APPLICATION NUMBER: US/10/994,726  
; PRIOR FILING DATE: 2004-11-23  
; PRIOR APPLICATION NUMBER: 09/830,230  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: PCT/US98/12718  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/057,483  
; PRIOR FILING DATE: 1997-09-03  
; PRIOR APPLICATION NUMBER: 60/053,344  
; PRIOR FILING DATE: 1997-07-22  
; PRIOR APPLICATION NUMBER: 60/053,377  
; PRIOR FILING DATE: 1997-07-22  
; PRIOR APPLICATION NUMBER: 60/050,359  
; PRIOR FILING DATE: 1997-06-20  
; NUMBER OF SEQ ID NOS: 756  
; SOFTWARE: Patent in Ver. 3.3  
; SEQ ID NO 322  
; LENGTH: 1087  
; TYPE: PRT  
; ORGANISM: Borrelia burgdorferi  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (54)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (594)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (703)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-994-726-322

Query Match 3.9%; Score 200; DB 18; Length 1087;  
Best Local Similarity 19.6%; Pred. No. 0.0021;  
Matches 212; Conservative 164; Mismatches 358; Indels 347; Gaps 54;  
QY 18 VTHQEVFLVKEPIKLTQASSISGADYAESGSKLKINETSGPVDDTVDLFSQKR 77  
Db 156 VNDQKNLENLEK--LKK-----NLGSKSENTLINDSQIE-----NDKQ 193  
QY 78 TT---PEKIKONLAGPREQE-----LKAVNTETSEKQITSGQLEQSK-----ES 121  
Db 194 NTNLSKEKNSNILKTPDNSKYSNNNTTSLKISSNSQKESLSPSPQTIIGKIYRPS 253  
QY 122 LSLNKTVPSTSNWEICDITKGTILVG-----LSKSGV-EKLSQTDHLVLPQAADGTQ 174

Db 254 YLIKKEL-----YELDDINTGRVTIGKRLKELIKGLSNKFKQKVNELIENSKKEASN 308  
QY 175 LI-----QVASFAPTPDKKTAIAEYTSRAGENGESQLDVGDKBIINEGEVFNYSY 224  
Db 309 LLTLTKIDIPNLINIPKDPYK-----EIFQLDKEDK-----POY 346  
QY 225 L-----LKKVTITPGYKHIOQDAFVD-----NKAIAEVLNPESELETI 261  
Db 347 LEDLKSQVHSIKPIDLENTKSR--QQAIDKOLNEFLKKNPNDAQASKTLAQANKIQHLEDL 404  
QY 262 SDYAFALHALKQIDLPNLK-----AIGELAF-----DNOITGKLSLPROLMLAERAF 311  
Db 405 KSKVH-----SIKPIDL-ENTKSRQQAIDKOLNEFLKKNPNDAQASKTLAQANKIQHLEDLKS 460  
QY 312 KSNHIKTIEF-----RGNLSKVICEASFDND-----LSQLMLPDGLEKIESEAF 357  
Db 461 KVHSIKPIDLENTKSRQQAIDKOLNE--FLKKNPNDAQASKTLAQANKIQHLEDLKS 518  
QY 358 GNPQDDHYNRVLVWTKSGKNPSGLATENTYVNDPKSLWQBSPEIDYTKWLEBFTYQKN 417  
Db 519 IKPID-----LENT-----KSRQQAIDKOLNE-----BFX-KNN 544  
QY 418 SVTGFSGKLGQKVRNKNLEIPKQHGVTITEIGDNAFRVDFONKTLRKYDLEEVKLP 477  
Db 545 PNDQAASKTLAQANKIQHLEDLKS-----VHSIKPIDLENTKSRQQAIDKOLNE-- 592  
QY 478 TIRKIGAFQFQSNLKSFEASDDLEIEKEGAFNNRIETLE-LKDKLVITIGDAAFHINHI 536  
Db 593 -----EPXKNPNDAQASKTLAQ-----NKIQHLEDLKSK-----VHSI 627  
QY 537 YAIVL-----PESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHDLSEBK 590  
Db 628 KPIDLENTKSRQQAIDKOL--NEFLKKNPNDAQASKTLAQ-----ANKIQHLEDLKSK 677  
QY 591 QLTEIPVQAFSDNALKEVLLPASLKTIREAFKKNHLKQLEVASALSHPALNDDNDGD 650  
Db 678 VHSIKPIDLENTKSRQ-----AIKOLNE--FXKNPNDAQASKTLAQ-AY-----ENNGD 725  
QY 651 ----EQFDNKVVVKT-----HHN-----SYALADGEHFI-----VDPDKLSS----- 683  
Db 726 LLKAENAYEKIILKLTNTQEDHYKLGIIIRPKLKKYEHSESFDQTIKLDPKHKKALHNKI 785  
QY 684 TIVDLKILKLIBGLDYSTLRQTQTFQDMTTAGKALLSKSNURQGEKQKFLQBAQ--- 740  
Db 786 ALMMLNKNKKAIESFE-----KAIQIDKNYGTAYYQKGAEEKNGD 826  
QY 741 -----PFLGRVDLDKAIKAERKALVTKKATNGQLLERSIN-----KAVLAYNNS 785  
Db 827 MQQAFASFKNAYNLDKNPNYALKAGIVNNLGNFKQSEYLNFPNNAKKNPEIATYNLS 886  
QY 786 AIKKANVRLKEKELDLTGLVE-----GKGPLAQATWQGVYLLKTPPL 828  
Db 887 IAKPEN-NKLEESLETINKAIDLNPESKSEVLYLKASINLKNKENYQNAISLYSLVIEKNP- 944  
QY 829 PLPEYVIGLVYFDKSGKLIYALD-MSDTIGEGQKDAYGN----- 867  
Db 945 ENTSAYINLAKAYEKSNGKSOAISTLEKIINKNNKALNNLILYKKEKNYQKAEIPEK 1004  
QY 868 PILNVDEDNEGYPHALAVATLADYEGDLITLNSKLSQTSIRQVPTAAYHRAGIFQAIQ 927  
Db 1005 AIIN--SDIEAKYNLATTLIEINDNTRAKOLLR-----EYTKLKPNNPEALHALGIIENY 1058  
QY 928 N 928  
Db 1059 N 1059

## RESULT 35

US-10-994-726-321  
; Sequence 321, Application US/10994726  
; Publication No. US20050147999A1  
; GENERAL INFORMATION:

```

; APPLICANT: Ni et al.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481D1
; CURRENT APPLICATION NUMBER: US/10/994,726
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: 09/830,230
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: Patent in Ver. 3.3
; SEQ ID NO 321
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (573)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (627)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (735)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-994-726-321

Query Match          3.9%; Score 200; DB 18; Length 1119;
Best Local Similarity 19.6%; Pred. No. 0.0022;
Matches 212; Conservative 164; Mismatches 358; Indels 347; Gaps 54;

QY 18 VTHQEVFLVKEPILKQTOASSISGADYAESGSKLKNINETSQPVDDTVTLFSDKR 77
DB 188 VNDQKNLFNLEK---LKK-----NLGKSNSEILNDSQKIB-----NDKQ 225
QY 78 TT---PEKIDKDLAKGPREQ-----LKAVENTESEKQITSGSOLEQSK-----PS 121
DB 226 NTNLSKERNSEILKTPDNSKYSNNNTTSLKKISSNQSELSPPSQTIIGKIYRPS 285
QY 122 LSLINKTPVSTSNWEICDFITKGNLTG-----LSKSGV-EKLSOTDHLVLPQAADGTO 174
DB 286 YLIKKEKEL-----YEILDDINTGRTVTLGKRLKELIKGLSNKFKQKVELIENSKKEASN 340
QY 175 LI-----QVASFAPTPDKKTAIABYTSRAGENGESIQLDVDGKEIINEGEVFNYSY 224
DB 341 LLLTLIKKIDIEPNLINIPDKYK-----EIFQDKEDKK-----PQY 378
QY 225 L-----LKKVITPTGKHIGQAFVD-----NKNIAEVLNPSLEPTI 261
DB 379 LEDLSKVHSIKPIDLENTKSR--QQAIKDLNEFLKNNPNDAQAASKTLAQANKIQHLEDL 436
QY 262 SDYAFHAHLKOIDLPDNLK-----AIGELAF-----DNOITGKLSLPRQLMRLAERAF 311
DB 437 KSKVH---SIKPIDL-ENTKSQQALIKOLNEFLKNNPNDAQAASKTLAQANKIQHLEDLS 492
QY 312 KSNHIKTIFF-----RGNLSKVIAGEASFQDND-----LSQLMLPDGLEKIESEAPT 357
DB 493 KVHSIKPIDLENTKSQQAIKOLNE--FLKNNPNDAQAASKTLAQANKIQHLEDLSKVHS 550
QY 358 GNPQDDHNNRVVLWTGSKNPSGLATENTYNNPDKSLWQESPEIDYTKWLEEDFTYQCN 417
DB 551 IKPID-----LENT-----KSRQQAIDKDLN-----EPX-KNN 576

418 SVTGFSGKLGQKVKRNKNLEIPKQHNQGVITTEIGNAFNRVDFQNKTRKYDLEEVKPLS 477
DB 577 PNDQAASKTLAQANKIQHLEDLSK-----VHSIKPIDLENTKSQQAIKDLN--- 624
QY 478 TIRKIGAFAPQSNULKSFEASDDLEIEKEGAFMNRRIETLE-LKOKLVITIGDAAFHINH 536
DB 625 -----EFXKNPNDAQAASKTLAQ-----NKIQHLEDLSK-----VHSI 659
QY 537 YAVLV-----PESVOEIGRSFRONGANNLIFMGSKVKTLCGEMAFLSNRLHDLSEOK 590
DB 660 KPIDLENTKSQQAIKDL--NEFLKNNPNDAQAASKTLAQ-----ANKIQHLEDLSK 709
QY 591 QUTEIPVQAFSDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSASHIAFNALDDNGD 650
DB 710 VHSIKPIDLENTKSQQ-----AIKDLNE--FXKNPNDAQAASKTLAQ--AY-----ENNCD 757
QY 651 -----EOPDNKVVVKT-----HHN-----SYALADGHEFI-----VDPDKLSS 683
DB 758 LLKAENAYEKIITNTQEDHYKLGIIIRFKLKKYEHSESFDQTTIKLDPKHKALHNKGI 817
QY 684 TIVDLEKILKLEGLDYSTLRQTTQTPRDMTTAGKALLSKNLSKQGEKQKFLQEAQ--- 740
DB 818 ALMMLNKKKAIESFE-----KAIQIDKNYGTAYYQKGIABEKNKD 858
QY 741 -----PFLGRVDDLKAIKAEKALVTTKATKNGQLLERSIN-----KAVLAYNNS 785
DB 859 MQQAFASFKNAVNLKPNYALKAGIVSNLGNFKQSEYLNFFNANAKKPEIAIYNLS 918
QY 786 AIKKANVRLEKEKELDLTLGLVE-----GKGPLAQATWQGVYLLKTP 828
DB 919 TAKFEN-NKLESLESTINKAIDLNPKESEYLYLKASINLKENYQNAISLYSLVIEKNP- 976
QY 829 PLPEYVIGLVVYFDKSGKLIYALD-MSDTIGBGQDAYCN----- 867
DB 977 ENTSYIINLAKAYESGKNSQSAISTELEKILNKNLNNLNLGILYKKNYOKAIEIFEK 1036
QY 868 FILNVEDNEGYHALAVATLADYEGDITKILNLSKLSQTSIRQVPTAAYHRAGIFQATQ 927
DB 1037 AIIN--SDIEAKYNLATTLEINDNTRAKDLR-----EYTKLPNNPEALHALGIIIEYN 1090
QY 928 N 928
DB 1091 N 1091

RESULT 36
US-09-932-183A-2
; Sequence 2, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
; APPLICANT: Betell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394CI-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-09-932-183A-2

Query Match          3.9%; Score 199.5; DB 9; Length 2285;
Best Local Similarity 20.9%; Pred. No. 0.0063;
Matches 215; Conservative 155; Mismatches 356; Indels 303; Gaps 54;
```

```
QY 81 EKIKNLAKPREQELKAVTENTES-EKQITSGSQLESKESLSN----- 125
Db 771 KKAQDDPEQS-QQTNEAITNKDSTDKLIQQYKELQKVKSRSLSDBEQEYLQVTOOL 829
QY 126 -KTVPSSTNWEICDPIITKGNLTVLGSKGVKLSQT-DHLVLPQAA-----DGTQ 174
Db 830 AQTFPAL----VKGYSQGNAILKTNKELEKAIENKTEYALKKQETRDSAKKTFEDASK 885
QY 175 LIQVASFAPDPKTAIAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTIPTG 234
Db 886 EIKSKDELQYKQ--IADYNDKGRPKWDLIADDDYKVAADKAK--QSMUKAQSDIESG 941
QY 235 YKHIGQDAFVDMKNIAEYNLPESLET-ISDYAFALHAKQIDLPNLKAIAGELAFFDNQI 293
Db 942 NAKVKDSVLSTANAYSSIDISNTLKTISD-----VVKLNKDKDLDLP-ELEKFSSSL 994
QY 294 TGKLSLPRQLMRLAERAFKSNHIKTIEFRGNSLKVIGEASPDNDLSQMLPDGLEKIES 353
Db 995 -GKLQEKMQ-----KALDSGDEKAFD--NAKK-----DLQSL----- 1024
QY 354 EAFGNPQDDHYNRNVLTWKGKPSGLATENTYVNPDKSLWQESPEIDYTKMLEEDPT 413
Db 1025 -----ETYSKSDSS-----IDVFKMSFD--K 1043
QY 414 YQKNSVTGFSNKGLOKVRKNKNLEIPKOHNGVTITEIGDNAPRVNVDQNKTLRKYDLEEV 473
Db 1044 AQKNIKDG--DKSLSSVKSEVG-----DLGETLAEAGNEA---EDFGK-----LKEA 1086
QY 474 KLPETIRKIGAFAPFOSNNLKSFEASDDELEETKEGAFMNNRIETLELKDCLVTIGDAAPHI 533
Db 1087 LDANSVDDIKAAIKEMSDAMQF--DSVDQVLNGDIFNN-----TKQVAPLND----- 1132
QY 534 NHIYAVLPESVQIGRSPQNGANNLI PMGSKVKTGEMAFLSNRLEHLDLSEFQKOLT 593
Db 1133 -----LLEKMAE-GKS-ISANEANTLI---QKDKELAQAIENGVVVKINRDEVIKQR 1180
QY 594 EIPVQAFSD-----NALK-----EVLPLPASLKTIRSE----- 620
Db 1181 KVKLDAYNDMVTYSNKLKMKTEVNNAITLNADTLRIDSLKLRKERKLDMSAEALSLEV 1240
QY 621 -----AFKQNHKOLE-----VASAL-SHIAFNALDNDGDDEQFD 654
Db 1241 KSINNVAADAKKELKLEKMLQPGGYSNSQIEAMQSVKSALESYI--SASEATSTQBMN 1298
QY 655 NKVVVTHNSYALADGEHFIVDPDKLSSTIVD-----LEKILKLEGL-----DYSTLR 704
```

```
RESULT 37
US-10-927-615-2
; Sequence 2, Application US/10927615
; Publication No. US20050009146A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394CL-US
; CURRENT APPLICATION NUMBER: US/10/927,615
; PRIOR FILING DATE: 2004-08-25
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-927-615-2
```

```
Query Match 3.9%; Score 199.5; DB 17; Length 2285;
Best Local Similarity 20.9%; Pred. No. 0.0063;
Matches 215; Conservative 155; Mismatches 356; Indels 303; Gaps 54;
QY 81 EKIKNLAKPREQELKAVTENTES-EKQITSGSQLESKESLSN----- 125
Db 771 KKAQDDPEQS-QQTNEAITNKDSTDKLIQQYKELQKVKSRSLSDBEQEYLQVTOOL 829
QY 126 -KTVPSSTNWEICDPIITKGNLTVLGSKGVKLSQT-DHLVLPQAA-----DGTQ 174
Db 830 AQTFPAL----VKGYSQGNAILKTNKELEKAIENKTEYALKKQETRDSAKKTFEDASK 885
QY 175 LIQVASFAPDPKTAIAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTIPTG 234
Db 886 EIKSKDELQYKQ--IADYNDKGRPKWDLIADDDYKVAADKAK--QSMUKAQSDIESG 941
QY 235 YKHIGQDAFVDMKNIAEYNLPESLET-ISDYAFALHAKQIDLPNLKAIAGELAFFDNQI 293
Db 942 NAKVKDSVLSTANAYSSIDISNTLKTISD-----VVKLNKDKDLDLP-ELEKFSSSL 994
QY 294 TGKLSLPRQLMRLAERAFKSNHIKTIEFRGNSLKVIGEASPDNDLSQMLPDGLEKIES 353
Db 995 -GKLQEKMQ-----KALDSGDEKAFD--NAKK-----DLQSL----- 1024
QY 354 EAFGNPQDDHYNRNVLTWKGKPSGLATENTYVNPDKSLWQESPEIDYTKMLEEDPT 413
Db 1025 -----ETYSKSDSS-----IDVFKMSFD--K 1043
```

```
QY 414 YQKNSVTGFSNKGLOKVRKNKNLEIPKOHNGVTITEIGDNAPRVNVDQNKTLRKYDLEEV 473
Db 1044 AQKNIKDG--DKSLSSVKSEVG-----DLGETLAEAGNEA---EDFGK-----LKEA 1086
QY 474 KLPETIRKIGAFAPFOSNNLKSFEASDDELEETKEGAFMNNRIETLELKDCLVTIGDAAPHI 533
Db 1087 LDANSVDDIKAAIKEMSDAMQF--DSVDQVLNGDIFNN-----TKQVAPLND----- 1132
QY 534 NHIYAVLPESVQIGRSPQNGANNLI PMGSKVKTGEMAFLSNRLEHLDLSEFQKOLT 593
Db 1133 -----LLEKMAE-GKS-ISANEANTLI---QKDKELAQAIENGVVVKINRDEVIKQR 1180
QY 594 EIPVQAFSD-----NALK-----EVLPLPASLKTIRSE----- 620
Db 1181 KVKLDAYNDMVTYSNKLKMKTEVNNAITLNADTLRIDSLKLRKERKLDMSAEALSLEV 1240
QY 621 -----AFKQNHKOLE-----VASAL-SHIAFNALDNDGDDEQFD 654
Db 1241 KSINNVAADAKKELKLEKMLQPGGYSNSQIEAMQSVKSALESYI--SASEATSTQBMN 1298
QY 655 NKVVVTHNSYALADGEHFIVDPDKLSSTIVD-----LEKILKLEGL-----DYSTLR 704
```

```
Db 1299 KOALVEAGTSLENWTDQOEKANEETKTSMYVVDKYKEALEKVNAEIDKYNKQVNDYPKYS 1358
QY 705 QTTOTQFRD-MTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLDAKAEKALVTKK 763
Db 1359 Q-----KYRDAIKKEIKALQOKKLMQ-EQAKLLQD-QIKSGNI-----TOYGIIVTST 1404
QY 764 ATKNG--QLLERSINKAVLAYNNSAIKKANVRKLEKELDLTGLVBEGK-GPLAQATWVG 820
Db 1405 TSSGGTPSTGSGYSKYSSYNSAASKYNV-----DPALIAAVIQESGFNAKARSVG 1459
QY 821 VYLLKTPPLPEYIYGLNVPDKSGKLIYALDMSDTIGEGQK-DAYGNPILNVDNEGY 879
Db 1460 ANGLMQLMPATAKSLGWNAYDP-----YQNVMGGTKYLAQOLEKFGG---NVEKALAAY 1511
QY 880 HALAVATLADYEGLD-----IKTIL---NSKLSOLTSIRQVPTAAVHRAGIFQAION 928
Db 1512 NA-GPGNVIKYGGIPPFKETQNYVKIMANYSKLSATS-----SIASY 1556
QY 929 AAAEAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG--SFY 986
Db 1557 -----TNNSAFRVSSKYQOESGLRSP--HKGTDFAAKAGTAIKSLQSGKV 1601
QY 987 GILGYTSTA 995
Db 1602 QIAGYSKTA 1610

RESULT 38
US-10-927-590-2
; Sequence 2, Application US/10927590
; Publication No. US2005059112A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394CI-US
; CURRENT APPLICATION NUMBER: US/10/927,590
; PRIOR FILING DATE: 2004-08-25
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PR1
; ORGANISM: Bacillus subtilis
US-10-927-590-2

Query Match 3.9%; Score 199.5; DB 17; Length 2285;
Best Local Similarity 20.9%; Pred. No. 0.0063;
Matches 215; Conservative 155; Mismatch 356; Indels 303; Gaps 54;

QY 81 EKIKNLAKPGREQLKAVNTES-EKOITSGSOLESKESLSLN----- 125
Db 771 KAKDDPEGS-QQTVEAITTNDKTDKLIQYKELQKVKERSLTSDEEQBYLQVTQOL 829
QY 126 -KTPVSTSNWEICDFTKGNLTVLGSKGVEKLSQT-DHLVLPQAA-----DGTQ 174
Db 830 AQTFPAL-----VKGYDSQGNALTKNKELEKAIENTKEYLALKQOTRDSAKKTFEDASK 885
QY 175 LIQVASFAFTPKKTAIATSRAGENGEISQLDVGKEIINEGEVFNISLLKVTIPTG 234
Db 886 EIKSKDELKQYKQ--IADYNDKGRPKWDLIADDDDDYKVAADKAK--QSMKLAQSDIESG 941
QY 235 YKHIGQDAFVDNKNTAEVNLPSLET-ISDYAFALHALKQIDLPNLKAIAGELAFEDNQI 293
Db 942 NAKVQDSVLSTANAYSSDISNLTLSISD-----VNMKFKDLDLP-EELKFPSSSL 994
QY 294 TCKLSLRLQMLRAERAFKSNHIKTIEFRGNSLKVIGBASFOQNDLSQLMLPDGLEKIES 353
```

```
Db 995 -GKLEQKMQ-----KALDSGDEKAPD--NAKK-----DLQSL----- 1024
QY 354 EAFITGNPGDDHNNRNVLTWTKSGKPSGLATENTYVNDPKSLQWQSPEDIDYTKWLEEDFT 413
Db 1025 -----ETYSKSDSS-----IDVFKMSFD--K 1043
QY 414 YOKNSVTGFSNKGLOKVRKNKLEIPKQHNGVTITTEIGDNAFRNVDFQNKTLRKVDLEEV 473
Db 1044 AQKNIKDG--DKSLSSVSEVG-----DLGETLAAGNEA--EDFGK-----LKEA 1086
QY 474 KLPSTIRKIGAPAFQSNLKLSPFASDDLEEEIKEGAFMNRRIETLEKOKLVITGDAAFHI 533
Db 1087 LDANSVDDIKKAIKEMSDAMQF--DSVQDVLNGDIFNN-----TKQVAPLND----- 1132
QY 534 NHIYAIVLPESESQETGRSAPRONGANNLIFMGSKVKTIGEMAFNLNLEHLDLSEKQILT 593
Db 1133 -----LLEKMAE-GKS-ISANEANTLI--OKDKELAQAISIENGVVVKINRDEVIKOR 1180
QY 594 BIPVQAFSD-----NALK-----EVLLPASLKTIRE----- 620
Db 1181 KVKLDAYNDMVTYSNKLAKMTEVNNNAIKTLNADTLRIDSILKLRKERKLDMSAEISLDEV 1240
QY 621 -----APKNHLKOLE-----VASAL-SHIAFNALDNDGDEQPD 654
Db 1241 KSINNVAADAKKELKLEKMLQPGGYSNSQIEAMQSVKSALESYI--SASEEATSTQEMN 1298
QY 655 NKVVVYKTHNSYALADGEHFIVDPDKLSSTIVD-----LEKLLKLEGL-----DYSTLR 704
Db 1299 KOALVEAGTSLENWTDQOEKANEETKTSMYVVDKYKEALEKVNAEIDKYNKQVNDYPKYS 1358
QY 705 QTTOTQFRD-MTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLDAKAEKALVTKK 763
Db 1359 Q-----KYRDAIKKEIKALQOKKLMQ-EQAKLLQD-QIKSGNI-----TOYGIIVTST 1404
QY 764 ATKNG--QLLERSINKAVLAYNNSAIKKANVRKLEKELDLTGLVBEGK-GPLAQATWVG 820
Db 1405 TSSGGTPSTGSGYSKYSSYNSAASKYNV-----DPALIAAVIQESGFNAKARSVG 1459
QY 821 VYLLKTPPLPEYIYGLNVPDKSGKLIYALDMSDTIGEGQK-DAYGNPILNVDNEGY 879
Db 1460 ANGLMQLMPATAKSLGWNAYDP-----YQNVMGGTKYLAQOLEKFGG---NVEKALAAY 1511
QY 880 HALAVATLADYEGLD-----IKTIL---NSKLSOLTSIRQVPTAAVHRAGIFQAION 928
Db 1512 NA-GPGNVIKYGGIPPFKETQNYVKIMANYSKLSATS-----SIASY 1556
QY 929 AAAEAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKG--SFY 986
Db 1557 -----TNNSAFRVSSKYQOESGLRSP--HKGTDFAAKAGTAIKSLQSGKV 1601
QY 987 GILGYTSTA 995
Db 1602 QIAGYSKTA 1610

RESULT 39
US-10-926-729-2
; Sequence 2, Application US/10926729
; Publication No. US20050106668A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394CI-US
; CURRENT APPLICATION NUMBER: US/10/926,729
; PRIOR FILING DATE: 2004-08-25
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
```







Db 2120 EVIKASPKVGPAPKVPCTPIKKEDQEVKRVKVKELP-NTGSEMDLPKLKALITGAALL 2178

RESULT 42  
US-11-020-509-2  
; Sequence 2, Application US/11020509  
; Publication No. US20050106648A1  
; GENERAL INFORMATION:  
; APPLICANT: FOSTER, Timothy et al.  
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .  
; FILE REFERENCE: P07263US02/BAS  
; CURRENT APPLICATION NUMBER: US/11/020,509  
; CURRENT FILING DATE: 2004-12-27  
; PRIOR APPLICATION NUMBER: US 10/172,502  
; PRIOR FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: US 60/298,098  
; PRIOR FILING DATE: 2001-06-15  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 2189  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-11-020-509-2

Query Match 3.9%; Score 199; DB 19; Length 2189;  
Best Local Similarity 19.7%; Pred. No. 0.0064;  
Matches 201; Conservative 176; Mismatches 359; Indels 284; Gaps 53;

Qy 16 SVVTHNQEVFSLVKEPILKQTQASSI-SGADYAESGSKLKNINETSPPVDDTVDLFS 74  
Db 1232 AINTLNQDROQAIES--IKQANTNAEVQAAATVAENN-----IDAVQDVVK 1276

Qy 75 DKRTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSKS-----SLSLNT--- 127  
Db 1277 -KQAAARDKITARVAK--RIEAVKQTPNATDEBKQ-AAVNQINQLKQAFNQINQNTNDQ 1332

Qy 128 VPSISNWEI-----CDFITKGNVLGSLGSGVEKLSQTDHLVLPQAAQDTQLIQVAS 180  
Db 1333 VDAATNQAINADNVAEAEVWIKPAIADIEKAVKEKQQQIDNSL----DSTNEK--EVAL 1387

Qy 181 FAFAPDKTAAEYTSRAGEIGEISQDVGKKEINEGEVFNLSLLKKVT--IPRGYKH 238  
Db 1388 QALAKEKEKALAA-IDQATNSQVNNQAATNGVSAIK-----IIQETKIKPAAREKI 1438

Qy 239 QDAFVDNKNATNAVLNPE-----SLETISDYFAFAHLAKQIDLPDLNLKAIGELAFF 289  
Db 1439 NOKA---NELRAQINQDKAEATAEERQAALDKIND-----LVAKAMTNITNDRNTNQVNS 1490

Qy 290 DNOITGKLSL--PROLMRAER-----APKSNHIKTIETFRGNSLKVIGASPODND-- 338  
Db 1491 TNQALDDIALVTPDHRIVRAAARDAVKQOYEAKKHEIEQAEHATDBEQVALNQLANNEK 1550

Qy 339 ----LSQMLPDGLKEISEAFTGNPGDDHY-----NNRVVLWTKSGKNP 379  
Db 1551 ALONINQAINNDVGRVENSIGIATUKGVEPHIVVPEAQEAIKASADNQV-----ESIKDT 1606

Qy 380 SGLATE-----NTYVNPDKLQWESPEIDYTKWLEEDFTYQKNSVTGFSNKLQKVK--- 431  
Db 1607 PHATTDELDEANQIN--DTLKQGGQDIDNT--TQDAAV--NDVRNQTIKAEIQKPKV 1659

Qy 432 -----RNKN-----LEIPKQNGVITIIEGD--NAFRNVDQNTKLRYD 469  
Db 1660 FRKRAALDNI DESNNQLDAIRNTLDUTQDERNVAIAALNKIVNAIKNDIAQNKTAEVD 1719

Qy 470 LEE-----VKL-----PSTIRKIGAFQSNLKSFEASD--DLREEIKEGA----- 508  
Db 1720 QTEADGNNNIKVILPKVQKPAARQSVSAKEAQNAL--IQOSDLSTEEERLAAKHLVEQ 1777

Qy 509 FMNNRIETLEKDKLVTTIGDAAFHINHIIYAVLP-----ESVQEIGRSAPRQNGAN 559  
Db 1778 ALNQAIDQINHADKTAQVQNQNSIDAQNIISKIPATTVKATALQOIQNIATNKINLIKAN 1837

Qy 560 N-----LIFMGSKVK---TLGEMAFEL-----SNRLBHL-----LSEQ 589  
Db 1838 NEATDEBQNAAIQVQEKELIKAKQOIAGAVTNADVAYLLHDKGRNEIREIEPIVINKKATAR 1897

Qy 590 KOLTEIPVQAFSDNALKEVLLPASIKTIREAFKKNHLKQLEVASALSASHIAFNALDNDG 649  
Db 1898 EQLTTL-----FND---KQAEANVOATVEE--RNSILAQLQ---NIYDTAIGQIDQRS 1945

Qy 650 DEQPDNKVVVKNTHNSYALADGE-HFIVDPDKLSSTIVDLEKILKLJEGLDYSTLRQTQ 708  
Db 1946 NAOVMDKTATL---NLQTIHLDVHPKPKDAEKTINDDLARVTHLVQ---NYRKVSDRNK 1999

Qy 709 TQPRDWTAGK-----ALLSKNLKQGEKQKFLQEAQFPLGRVD----- 747  
Db 2000 ADALKAITALKLQMDDEELKTARTNADVDVLRKFNVALGDIETAVITEKENSLLRIDNTAQ 2059

Qy 748 ----LDKAIKAAE-----KALVTKKATKNGQLLERS-----INKAVLAYNNSA 786  
Db 2060 QTYAKFKALATPEQLAKVKKALIDQYVADGNRMVDEATLNDIKKDTQLIIDELIAIKLPA 2119

Qy 787 -----IKKAN---VKRLEKELDLLTGLVEGKGPLAQATMVQGVYLL 824  
Db 2120 EVIKASPKVGPAPKVPCTPIKKEDQEVKRVKVKELP-NTGSEMDLPKLKALITGAALL 2178

RESULT 43  
US-10-732-923-3331  
; Sequence 3331, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 3331  
; LENGTH: 1790  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-732-923-3331

Query Match 3.9%; Score 198.5; DB 17; Length 1790;  
Best Local Similarity 21.1%; Pred. No. 0.0052;  
Matches 249; Conservative 173; Mismatches 425; Indels 333; Gaps 59;

Qy 25 FSLVKEPILKQTA--SSSISGADYAESGSKLKNINETSPPVDDTVD----- 71  
Db 434 FDLQDFLLKQVQLCNNSTNNVGNNAKENGSGNSKSDKSDKDTGDKGTEYEGSPKAN 493

Qy 72 LF-----SDKRTTPEKI-----KONLAKGPREQELKAVTENTSEKQITSG 112  
Db 494 LFEVLLNYDAELNLNPKLPFTTIDIFMFFQODH-----KYSEELREITRN-----VTTG 543

Qy 113 SQLEQSKESLSLNTVP---STS-----NWEICDFTTKGNTLVGLSKS 152  
Db 544 NDLE-DEEPLKAIQITISELLTSLTAADIRIPISYLTYFLIYWLFGDFKATNDFL--SDKS 600

Qy 153 GVEKLSQTDHLVLPQAAQ-----GTQLIQVA---SFAFTPKKTAIAEY-TSRAGEN- 201  
Db 601 VIKSLLSFSYQI---QDEVDVTKLVTMLLGVAVEFSKSPFKPKSYFFETITLTKGN 657

Qy 202 --GISQLDQVD---GKEIINEGEVFNLSL-----LKKVTIPTGYKHIQDAFV----- 244  
Db 658 YASRIKQPKKDSYKSDVMNEDSILTELDGTGLPKVYFSTYFIQLFNENIYRIRITALSH 717

Qy 245 --DNKNIAEVLNPE-----SLETISDYFAFAHLAKQIDLPDLNLKAIGEL 286  
Db 718 DPDEEPIKISFEBVEKLRQCKTKGKEITSLQETESTHENLTKELIALTNEHKELDEK 777

Qy 287 AFFDN-----QITGKLSLIPQLMRLAERAFKSNHIKTIETFRGNSLKV 328

Db 778 YQILNSHSSSLKENFISILETELKKNVRDSDLEMTQLRDVLETKDKNQALLEVYKSTIHK- 836  
QY 329 IGEASFDNDLSQMLPDGLKEIESEAFNGPDHYNRVVLTWSGKNPSGLATENTY 388  
Db 837 -----QEDSIK--TLEKLETLISQK--KKAEDGIN-----KMGKDLFALSREMOA 878  
QY 389 VNPD-KSLWQESPEIDY-----TKWLEEDFTYQKNSVTGFSNKGLOKVK-RNKNLEIPKQ 441  
Db 879 VEENCKNLQKXKSNVNHQKETSUKEDIAKITEIKAI-NENLEEMKIQCNNLSKEKE 937  
QY 442 HNGVTITEIGD-----NAFRNVDFQNTLRKYDLEEVKLPESTIR- 480  
Db 938 HISKELVEYKSRFQSHDNLVAKLTEKLKSLANNYKDMQAEENSLIK-AVEESKNESLIQ 996  
QY 481 -----KIGAPAFQSNLKK-----SFEAS-----DDLEEIKEGAFMNNRIETLE----- 518  
Db 997 SNLQNKIDMSQEKENFOIERGSEIEKNIQKLTISDLEQTKETEEIISKSDSSKDEYESOI 1056  
QY 519 --LKDKLVITIGDA-AFHINHIAIVLPESVQIGRSAPRQNGANNL-IFMGSKVKTGEM 574  
Db 1057 SLUKEKLEATTANDENVNKSILTKTREETEAEALAAVK-NLKNLETKLETSEKALKEV 1115  
QY 575 AFLSNRLEHLDLSEQKLTETIPVQAFSDNALKEVL-----LPASLKTIRREAFKKNHLK 628  
Db 1116 KENEHLKEEKIQLEKEATETKQQLNSLRANLESLEKEHEDLAAQLKYEEQIANKEQY 1175  
QY 629 QLEVASALSHIAFNALDNDGDEQPDNVVTHNSYALADGEHFIV----- 676  
Db 1176 NEEISQ-----LNDEITSTQENESIKKKND-----LEGEVKAMKSTSEOSNLKK 1222  
QY 677 -DPDKLSSTIVDLEK-----ILKLEGLDYSTL-----ROTTQTFQDRMTT 716  
Db 1223 SEIDALNLQIKELKKNETNEASLESIKSVESETVKIKELQDECNFKEVSELEDK 1282  
QY 717 AGKALLSKNLQKQKQKFLQ-----EAQFFLGRV-DLDKAIKAEKALVTKKAT--- 765  
Db 1283 ASEDKNSKYLELQKSEKIEKELDAKTTELKIQLEKITNLSKAKESESELRLKTSSE 1342  
QY 766 --KNG-QLLERSINKAVLAYNNSAIVKAVKELEKELDLTLGLVSGKPLAQATVQGVY 822  
Db 1343 ERKNAEQLEKUKNE--IQIKNOA-----FEKERKLN--EGSSTITO----- 1381  
QY 823 LLKTLPLPEYIYGLNVYDPKSGKL-----IYALDMSDTIGRGQDAYGNPILANVDEDE 877  
Db 1382 -----EYSEKINTLEDELIRLQENELAKAEIDNTRSELEKVSINDELSEKQN- 1431  
QY 878 GYHALAVATLADYEGLDIKTILNSKLSQTSI-----ROVPTAAYHRAGIFOAIQNAAB 932  
Db 1432 -----TIKSLOD-EILSVKDKITRDNDEKLLSTIERDNKRDLESLEKQ-----LRAAQESKAK 1481  
QY 933 ABQLLPKPTHTSEKSSSESA--NSKD--RGLOSNPKTN 967  
Db 1482 VEGLUK-----LEEESSEKAELEKSKEMMKLESTIESN 1517

## RESULT 44

US-10-369-493-1586  
; Sequence 1586, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 1586  
; LENGTH: 1790  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-1586

Query Match 3.9%; Score 197.5; DB 15; Length 1790;  
Best Local Similarity 21.0%; Pred No. 0.006;  
Matches 248; Conservative 174; Mismatches 425; Indels 333; Gaps 59;

QY 25 FSLVKEPILKQPA--SSSISGADYAESGSKJAKINETSQVDDTVTD----- 71  
Db 434 FDLQDFLLKQVLCNNSTNNVGNKENGSGSKSDKSDKDTGDKGTVEGSEFKAN 493  
QY 72 LF-----SDKRTTPEKI-----KDNLAGPREQELKAVTENTSEKQITSG 112  
Db 494 LFEVLNVDALNLPFKLFTTDFIMFFFOQDH-----KYSELREITRN-----VTG 543  
QY 113 SOLEQSKESLSLNTKTPV--STS-----NWEICDFTKGMTLVGLSKS 152  
Db 544 NDLE-DEEPLKAIQIISLTTSLTAADIRIPISVLTLYWLFQGFKAINDFL--SDKS 600  
QY 153 GVEKLSQTDHLVLPQAAD-----GTOLIOVA--SPAFTDPKKTATAEY--TSRAGEN- 201  
Db 601 VIKSLLSFSYQI--QDEEDVTIKCLVTMLLGVAEYFSSKESPPRKEYPEFITKTLGKON 657  
QY 202 --GEISQLDVD--CKEIIINEGEV-----FNSYLLKKVT-----IPTGYKH 237  
Db 658 YASRIKQPKQDSYFSGVDMNEDSILTPELDEGLPKVYFSTYFIQLFNENIYRIIRALSH 717  
QY 238 IQQDAFVNKNIAEYNLPE-----SLETISDYAFALHAKQLDLPDLNLAIGEL 286  
Db 718 DPDEPINKISFEVEKLEQROCTKJKEITSQTESTHENTKELIALTWEHKEDEK 777  
QY 287 APFDN-----QITGKLSLPRQLMLRAEAFKSNHNTIIFRGNLSKV 328  
Db 778 YQILNSHSSSLKENFISILETELKKNVRDSDLEMTQLRDVLETKDKNQALLEVYKSTIHK- 836  
QY 329 IGEASFDNDLSQMLPDGLKEIESEAFNGPDHYNRVVLTWSGKNPSGLATENTY 388  
Db 837 -----QEDSIK--TLEKLETLISQK--KKAEDGIN-----KMGKDLFALSREMOA 878  
QY 389 VNPD-KSLWQESPEIDY-----TKWLEEDFTYQKNSVTGFSNKGLOKVK-RNKNLEIPKQ 441  
Db 879 VEENCKNLQKXKSNVNHQKETSUKEDIAKITEIKAI-NENLEEMKIQCNNLSKEKE 937  
QY 442 HNGVTITEIGD-----NAFRNVDFQNTLRKYDLEEVKLPESTIR- 480  
Db 938 HISKELVEYKSRFQSHDNLVAKLTEKLKSLANNYKDMQAEENSLIK-AVEESKNESLIQ 996  
QY 481 -----KIGAPAFQSNLKK-----SFEAS-----DDLEEIKEGAFMNNRIETLE----- 518  
Db 997 SNLQNKIDMSQEKENFOIERGSEIEKNIQKLTISDLEQTKETEEIISKSDSSKDEYESOI 1056  
QY 519 --LKDKLVITIGDA-AFHINHIAIVLPESVQIGRSAPRQNGANNL-IFMGSKVKTGEM 574  
Db 1057 SLUKEKLEATTANDENVNKSILTKTREETEAEALAAVK-NLKNLETKLETSEKALKEV 1115  
QY 575 AFLSNRLEHLDLSEQKLTETIPVQAFSDNALKEVL-----LPASLKTIRREAFKKNHLK 628  
Db 1116 KENEHLKEEKIQLEKEATETKQQLNSLRANLESLEKEHEDLAAQLKYEEQIANKEQY 1175  
QY 629 QLEVASALSHIAFNALDNDGDEQPDNVVTHNSYALADGEHFIV----- 676  
Db 1176 NEEISQ-----LNDEITSTQENESIKKKND-----LEGEVKAMKSTSEOSNLKK 1222  
QY 677 -DPDKLSSTIVDLEK-----ILKLEGLDYSTL-----ROTTQTFQDRMTT 716  
Db 1223 SEIDALNLQIKELKKNETNEASLESIKSVESETVKIKELQDECNFKEVSELEDK 1282  
QY 717 AGKALLSKNLQKQKQKFLQ-----EAQFFLGRV-DLDKAIKAEKALVTKKAT--- 765

Db 1283 ASKDKNKYELQKSEKIKEELDAKTTELKIQLEKITNLKAKEKSELSRLKKTSS 1342  
Qy 766 --KNG-QLLERSINKAVLAYNSAIKANVRLEKELDLTLGLVEGKPLAQATWQGVY 822  
Db 1343 ERKNAEQLEKLNQNE--IQKNQA-----FEKERKLN-----EGSSTITQ----- 1381  
Qy 823 LLKTPPLPEYVIGLVNVPFGSKL-----IVALDMSDTIGEGOKDAVGNPILNVDEDE 877  
Db 1382 -----EYSEKINTLEDELIRQENELKAEIDNTRSELEKVSLSNDELLEEKQN- 1431  
Qy 878 GYHALAVATLADYEGDLIKTLNSKLSQTSI-----RQVPTAAYHRAGIFQAIQNAAE 932  
Db 1432 -----TIKSLQD-EILSYKDKITRNDKLSIERDNKEDLSLKEQ-----LRAAQESKAK 1481  
Qy 933 AEQLLPKPGTHSEKSSSSESA-----NSKD--RGLQSNPKTN 967  
Db 1482 VEEGLK-----LEBESSKEAELEKSKEMMKKLESTIESN 1517

RESULT 45  
US-10-732-3330  
; Sequence 3330, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 3330  
; LENGTH: 1790  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-732-3330

Query Match 3.9%; Score 197.5; DB 17; Length 1790;  
Best Local Similarity 21.0%; Pred. No. 0.006;  
Matches 248; Conservative 174; Mismatches 425; Indels 333; Gaps 59;

Qy 25 FSLVKEPIKQTOA--SSSISGADYAESGSKLKNINSGPVDVTVD----- 71  
Db 434 FDLQRDFLLKQVLCNNTNNVGDNAKNGSKSDKSDSDKDTGDKDGTGYEGSKAN 493  
Qy 72 LF-----SDKRTTPEKI-----KDLAKGPREQELKAVTENTSEKQITS 112  
Db 494 LFEVLNLYDAELNLPFLFFTTDIFMFFQQDH-----KYSELEIRTN-----VTG 543  
Qy 113 SOLQSKESLSLNTKTPV---STS-----NWEICDFTKGNLTGLSLKS 152  
Db 544 NDLE-DEEPLKAIQITISELLTSLTAADIRIPISYLTFLIYVWFGDFKATNDFL--SDKS 600  
Qy 153 GVEKLSQTDHLVPSQAD-----GTQLIQA-----SFAFTPKKTAIAEY-TSPAGEN- 201  
Db 601 VIKLSLSFSYQI--QDESDVTIKCLVTMLGVAYEFSSKESFPKRYEFITKTLGKDN 657  
Qy 202 --GEISOLDVD--GKEIINEGEV-----FNSYLLKKVT-----IPTGYKH 237  
Db 658 YASRIKQPKDSYFKVDNEDSILTPDELGLPKVYFSTYFIQLFENIYRIRITALSH 717  
Qy 238 IQQDAFVNKNIAVNLPE-----SLETISDYAFAHALKQIDLDPNLKAI 286  
Db 718 DPDEEPINKISPEEVEKLRQCTKLKGEITSIQTESHTENLTKLIATNHEKELDEK 777  
Qy 287 AFPDN-----QITGKLSLPRQLMRLAERAFKNHKTTEFRGNSLKV 328  
Db 778 YQLNSSHSLKENFSILETELKNYRDSLDEMTQLRDVLETKDKENQTALLEYSKTIHK- 836  
Qy 329 IGCAEFQNDLSQLPDGLEKIESEAFGNPGDDHNNRVVLWTKSGNPGSLATENTY 388  
Db 837 -----QEDSIK--TLEKGLTILSQK-----KKAEDGIN-----KMGKOLFALSREMQA 878

Qy 389 VNPD-KSLWOESPEIDY-----TKWLEEDFTYQKNSVTGFSNKGLOKVK-RNKNLEIPKO 441  
Db 879 VEENCKNLQKEKOKSNVNHQKETSLSKABDIAAKITEIKAI-NENLEEMKIQCNLSKEKE 937  
Qy 442 HNGVTITEIGD-----NAFRNVDFQNKTLRKLYDLBEVKLPSTIR- 480  
Db 938 HISKELVEYKSRFQSHDNLVAKUTEKLSLANNYKQMAENESLIK-AVESKNNESSIQL 996  
Qy 481 -----KITGAFAPQSNNUK-----SFEAS-----DDLEIKEGAFNWNRIETLE----- 518  
Db 997 SNLQNKIDMSQKQENFQIERGSIETKIEQIKKTISDLQETKEBIIISKSPSSODEYESQI 1056  
Qy 519 --LKQKLVITIGDA-AFHINHIYAVLPESVQIEGRSAFRONGANL-LFPGSKVKTTGEM 574  
Db 1057 SLLKBELETATTANDENVNKISELTKTREBLEAELAAAYK-NLKNELETKLETSEKALKEV 1115  
Qy 575 AFLSNRLEHLDLSEKQLTPIPVQAFSDNALKEVL-----LPASLKTIREEAFKQNHK 628  
Db 1116 KENEHLEKKEKIQLEKEATETKQOLNSLRANLSLEKEHEDLAAQLKYEIQIANKEROY 1175  
Qy 629 QLEVASALSHIAFNALDNDGDEQFDMKVVVVKTTHNSYALADGHEFIV-----RQTOTQFRDMYT 716  
Db 1223 SEIDALMLQIKELKKKQNETNEASILLESIKSVESSTVIKELQDCNPKFKEVSELDKJK 1282  
Qy 717 AGKALLSKNLROGEKOKFLQ-----EAQFFLGRV-DLDKAIKAEKALVTKKAT--- 765  
Db 1283 ASEDKNKYELQKSEKIKEELDAKTTELKIQLEKITNLKAKEKSELSRLKKTSS 1342  
Qy 766 --KNG-QLLERSINKAVLAYNSAIKANVRLEKELDLTLGLVEGKPLAQATWQGVY 822  
Db 1343 ERKNAEQLEKLNQNE--IQKNQA-----FEKERKLN-----EGSSTITQ----- 1381  
Qy 823 LLKTPPLPEYVIGLVNVPFGSKL-----IVALDMSDTIGEGOKDAVGNPILNVDEDE 877  
Db 1382 -----EYSEKINTLEDELIRQENELKAEIDNTRSELEKVSLSNDELLEEKQN- 1431  
Qy 878 GYHALAVATLADYEGDLIKTLNSKLSQTSI-----RQVPTAAYHRAGIFQAIQNAAE 932  
Db 1432 -----TIKSLQD-EILSYKDKITRNDKLSIERDNKEDLSLKEQ-----LRAAQESKAK 1481  
Qy 933 AEQLLPKPGTHSEKSSSSESA-----NSKD--RGLQSNPKTN 967  
Db 1482 VEEGLK-----LEBESSKEAELEKSKEMMKKLESTIESN 1517

RESULT 46  
US-09-815-242-5635  
; Sequence 5635, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727

```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5635

Query Match      3.9%; Score 197.5; DB 9; Length 2368;
Best Local Similarity 19.8%; Pred. No. 0.0088;
Matches 242; Conservative 165; Mismatches 459; Indels 359; Gaps 59;

QY 19 THNOEVSFLKVPILKQTOASSISGADYAEBSG-----KSKLKINETSGPVDVTDVTD 71
Db 19 THNOEVSFLKVPILKQTOASSISGADYAEBSG-----KSKLKINETSGPVDVTDVTD 71
QY 72 LFSDKRTTPEKID-NLAKGPREQELKAVTENTESEKQITSGSQLESKESLSLNTKTVPS 130
Db 72 LFSDKRTTPEKID-NLAKGPREQELKAVTENTESEKQITSGSQLESKESLSLNTKTVPS 130
QY 904 IQTOAKVKAADTEVENAYNTRKQEIQNSNASTTEBKQ-AAYTELDTKKQEARLNDAAN 962
Db 904 IQTOAKVKAADTEVENAYNTRKQEIQNSNASTTEBKQ-AAYTELDTKKQEARLNDAAN 962
QY 131 TSNWEICDPIT-KGNTLVGLS-----KSGVVKLS----- 158
Db 131 TSNWEICDPIT-KGNTLVGLS-----KSGVVKLS----- 158
QY 963 TNS-----DVTTKADNGIAINQVQAATTKSKDAKAEIAQKASERKTAIEAMNDSTTEEQ 1018
Db 963 TNS-----DVTTKADNGIAINQVQAATTKSKDAKAEIAQKASERKTAIEAMNDSTTEEQ 1018
QY 159 -----OTDHLVLPQA-----ADGTLQIVASF-AFTPD-----KKTAL-----A 192
Db 159 -----OTDHLVLPQA-----ADGTLQIVASF-AFTPD-----KKTAL-----A 192
QY 1019 AAKDKVDQAVVTANADIDNAAANTVDNNAKTNEATIAAITPDANVPKAKAIADKQVA 1078
Db 1019 AAKDKVDQAVVTANADIDNAAANTVDNNAKTNEATIAAITPDANVPKAKAIADKQVA 1078
QY 193 EYTSRAGENGESQLDNDGKEIINEGEVFNYSLLKKVITPTGYKHGQDAFVD---NKN 249
Db 193 EYTSRAGENGESQLDNDGKEIINEGEVFNYSLLKKVITPTGYKHGQDAFVD---NKN 249
QY 1079 QETATDANNATTEKAAKQVQ-----TEKTTADTAIDGAHTNAEVEAAKNAEI 1129
Db 1079 QETATDANNATTEKAAKQVQ-----TEKTTADTAIDGAHTNAEVEAAKNAEI 1129
QY 250 AEVNPESLETSDYAFALHAKQLIDPNLKAIGEL-----AFPDNQTGKLSLP 300
Db 250 AEVNPESLETSDYAFALHAKQLIDPNLKAIGEL-----AFPDNQTGKLSLP 300
QY 1130 AKIEATQPATTTKDNKAQAIATKANERKTAIAQTODITAEIAAANANVDNAV----- 1183
Db 1130 AKIEATQPATTTKDNKAQAIATKANERKTAIAQTODITAEIAAANANVDNAV----- 1183
QY 301 ROLMLRAERAPKSNHKTIEFRGNSLKVIGESFQDNDLSQLMLPDGLEKIESEAFNG- 359
Db 301 ROLMLRAERAPKSNHKTIEFRGNSLKVIGESFQDNDLSQLMLPDGLEKIESEAFNG- 359
QY 1184 -QANNIEAANSQNDV-----DOAKTTGEASIDQ-----VTPTVNNKATAVTDAXNN 1229
Db 1184 -QANNIEAANSQNDV-----DOAKTTGEASIDQ-----VTPTVNNKATAVTDAXNN 1229
QY 360 ---PGDDHYNNRVLMWTKSGKN-----PSGLATENTYVNPDKSLWQESPEIDYTKWLEE 410
Db 360 ---PGDDHYNNRVLMWTKSGKN-----PSGLATENTYVNPDKSLWQESPEIDYTKWLEE 410
QY 1230 ITAATDD---NGVDYAKDAGKNSIQSTQPATVKSNAKNDVDQAVTTQQAIDNTTGAT- 1285
Db 1230 ITAATDD---NGVDYAKDAGKNSIQSTQPATVKSNAKNDVDQAVTTQQAIDNTTGAT- 1285
QY 411 DFTYQKNSVTGSPNGKLOKVKRNKLEIPKQNGVITIEIGNAFNV-----DFQNTKL 465
Db 411 DFTYQKNSVTGSPNGKLOKVKRNKLEIPKQNGVITIEIGNAFNV-----DFQNTKL 465
QY 1286 ---TEEKNAKDEL-----VLKAKEKAYQDILNAQTNDVDTQIKQAVADVQGITADTTIKDV 1339
Db 1286 ---TEEKNAKDEL-----VLKAKEKAYQDILNAQTNDVDTQIKQAVADVQGITADTTIKDV 1339
QY 466 RKVDL-----BEVKLPSTI-----RKIGAFAPQSN-NLKSFEASDDEIEKEG 507
Db 466 RKVDL-----BEVKLPSTI-----RKIGAFAPQSN-NLKSFEASDDEIEKEG 507
QY 1340 AKDELATKAREKALIAQTADATTEKEQANQOVDAELTQGNQNIENASQISDDVNTAKDN 1399
Db 1340 AKDELATKAREKALIAQTADATTEKEQANQOVDAELTQGNQNIENASQISDDVNTAKDN 1399
QY 508 AFM-----NNRIETL-ELKOKLVITIGDAAAFHINHIYAIVLPESVQIEG--RS 551
Db 508 AFM-----NNRIETL-ELKOKLVITIGDAAAFHINHIYAIVLPESVQIEG--RS 551
QY 1400 AIGADIPQASTDVKTNARAEILLTEMQWIKITILNNETTNE-----EKGNDIGPVRA 1452
Db 1400 AIGADIPQASTDVKTNARAEILLTEMQWIKITILNNETTNE-----EKGNDIGPVRA 1452
QY 552 AFRQNGANNLIIPMGSKVTKLGEMAFLSNRLHLDLSEKQLTEIPVQAFSDNALKEVLLP 611
Db 552 AFRQNGANNLIIPMGSKVTKLGEMAFLSNRLHLDLSEKQLTEIPVQAFSDNALKEVLLP 611
QY 1453 AY-EGLNMI---NAATTG-----DVTAKDTAVQKQQLHANPVKK---P 1492
Db 1453 AY-EGLNMI---NAATTG-----DVTAKDTAVQKQQLHANPVKK---P 1492
QY 612 ASLKTIREFAP--KKHLKQL-----EVASALSHIAFNALDNDGDGFQFNK 656
Db 612 ASLKTIREFAP--KKHLKQL-----EVASALSHIAFNALDNDGDGFQFNK 656
QY 1493 AG-KTALDQAAADKTKTIEQTPNASQOEINDAKQEVDTLNOAKTN-IDQSSTDEYVDNA 1550
Db 1493 AG-KTALDQAAADKTKTIEQTPNASQOEINDAKQEVDTLNOAKTN-IDQSSTDEYVDNA 1550
```

```
QY 657 V-----VVKRT--HHNSYALADGEHF-----IVDPD-----KLSSTIVDL-EKILKLEIG 697
Db 657 V-----VVKRT--HHNSYALADGEHF-----IVDPD-----KLSSTIVDL-EKILKLEIG 697
QY 1551 VKEGKAKINAVKTFSEYKKDALAKIEAAYNAKAVTEADNSNASTSSIEAEAKQKLAELKOT 1610
Db 1551 VKEGKAKINAVKTFSEYKKDALAKIEAAYNAKAVTEADNSNASTSSIEAEAKQKLAELKOT 1610
QY 698 LDYSTLRQTTQTFQFRDMMTAGKALLSKNSLKQGEKQKFLQEAQFPL-----GRVDLDKATA 753
Db 698 LDYSTLRQTTQTFQFRDMMTAGKALLSKNSLKQGEKQKFLQEAQFPL-----GRVDLDKATA 753
QY 1611 ADQNVNQATSKDDIEVQIHNDLDINDYTIPTGKESATTDILYAYADQKKNNISADTNAT 1670
Db 1611 ADQNVNQATSKDDIEVQIHNDLDINDYTIPTGKESATTDILYAYADQKKNNISADTNAT 1670
QY 754 KAEKALVTKKATKNGQLLERSINKAV-----LAYNNSAI-----KKAN----- 791
Db 754 KAEKALVTKKATKNGQLLERSINKAV-----LAYNNSAI-----KKAN----- 791
QY 1671 QDEKQQAIAKQVQNDQVQTALESINNGVDNGVDVDDALTOGKAALDAIQVDAIVKPKANQVID 1730
Db 1671 QDEKQQAIAKQVQNDQVQTALESINNGVDNGVDVDDALTOGKAALDAIQVDAIVKPKANQVID 1730
QY 792 -----VKRLEKELDLTLTG-----LVEGKGFLAQATMVQGVYLLKTLPLPLPEY 833
Db 792 -----VKRLEKELDLTLTG-----LVEGKGFLAQATMVQGVYLLKTLPLPLPEY 833
QY 1731 AKABETKESIDQSDQLTAEKTEALAMIKQITDQAKQGITDATTTAEEVEKAKAQ----- 1784
Db 1731 AKABETKESIDQSDQLTAEKTEALAMIKQITDQAKQGITDATTTAEEVEKAKAQ----- 1784
QY 834 YIGL-----NVYFDKSGK-----LIYALDM-----SDTIGEGQKQAYGNPILNVDBD 875
Db 834 YIGL-----NVYFDKSGK-----LIYALDM-----SDTIGEGQKQAYGNPILNVDBD 875
QY 1785 --GLEAFDNIQIDSTEKQKAIIELETAIDQIEAGVNVVDADATTE-EKEAFTNAL----- 1835
Db 1785 --GLEAFDNIQIDSTEKQKAIIELETAIDQIEAGVNVVDADATTE-EKEAFTNAL----- 1835
QY 876 NEGHALAVATLADY-EGLDIKTILNSKLSQTSIRQVPTAAYHRAGI-----FQAI 926
Db 876 NEGHALAVATLADY-EGLDIKTILNSKLSQTSIRQVPTAAYHRAGI-----FQAI 926
QY 1836 -EDILSKATEDISDQTTNABEATVKNLSALEQKQRIINPVVKNQNALEAIREVVNNQIEII 1894
Db 1836 -EDILSKATEDISDQTTNABEATVKNLSALEQKQRIINPVVKNQNALEAIREVVNNQIEII 1894
QY 927 QNAABAE-----QLLPKPGTHSEKS-----SSSESAN 954
Db 927 QNAABAE-----QLLPKPGTHSEKS-----SSSESAN 954
QY 1895 KNADADASAKETARTDLGRYDFRFPADKLDKTKTNTNEVAELQNVVTIPATEAIVPQNDPNAN 1954
Db 1895 KNADADASAKETARTDLGRYDFRFPADKLDKTKTNTNEVAELQNVVTIPATEAIVPQNDPNAN 1954
QY 955 SKDRGLQSNPKTNRGHSAIILPRTG 979
Db 955 SKDRGLQSNPKTNRGHSAIILPRTG 979
QY 1955 DTNSGSDNNDATANSNANATPENTG 1979
Db 1955 DTNSGSDNNDATANSNANATPENTG 1979
```

## RESULT 47

```
US-09-815-242-12389
; Sequence 12389, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12389
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12389
```

Query Match 3.9%; Score 197.5; DB 9; Length 2368;  
Best Local Similarity 19.8%; Pred. No. 0.0088;  
Matches 242; Conservative 165; Mismatches 459; Indels 359; Gaps 59;

QY 19 THNQVFSLVKEPIIKQTOQASSISSGADYABSSG-----KSKLKINETSPPVDDTVD 71  
DB 850 TONATVSNATNEEVAEADA-----VEAQKQGLHDIQVVKSKQEVADTKSVLKDINA 903  
QY 72 LPSDKRTTPEKID-NLAKGPREQELKAVTENTSEKQITSGSLQBSKESLSLNTKTVPS 130  
DB 904 IQTQAKVPAADTEVENAYNTRKQEIQNSASTTEBKQ-AAYTELDTKKQEARNLDAAN 962  
QY 131 TSNWEICDPIT-KGNTLVGLS-----KSGVEKLS----- 158  
DB 963 TNS-----DVTAKONGIAINQVQAATTKKSDAKAEIAQKASERKTAIEAMNDSTTESQ 1018  
QY 159 -----QTDHLVLPQQA-----ADGTQLIQVASF-AFTPD-----KKTAI-----A 192  
DB 1019 AAKQKVDQAVTANADIDNAANTDNDVNAKTTEATIAAITPDANVKPTAKQAIADKVOA 1078  
QY 193 EYTSRAGENGESQLDVGKEIINEGEVFNVSLLKKVTIPTGYKHIGQDAFVD---NKNI 249  
DB 1079 QETAIDANNGATTEKAAKQOVQ-----TEKTTADTAIDGAHTNAEVEAAKNAEI 1129  
QY 250 AEVNLPSLETISDYAFALHALKQIDLPNLKAIGEL-----APFDNQITGKLSLP 300  
DB 1130 AKIEAIQPATTKONAKQAIATKANERKTAIAQTODITAEIAAANAVNDNAV----- 1183  
QY 301 RQMLRAERAPKSNHIKTIIEFRGNSLKVIGASFOQNDLSQMLPDGLEKIESEAFQCN- 359  
DB 1184 -QANNIEAANSQNV-----DQAKTTGEASIDQ-----VTPVVKKATAVTDACKN 1229  
QY 360 ---PODDHNNRVVLWTSKGN-----PSGLATENTYVNPDKSLWQSPSEIDYTKWLB 410  
DB 1230 ITAATDD---NGVDATKADAGKNSIQSTQPATVAKSNKAKNDVDQAVTTQNTQAIQNTGAT- 1285  
QY 411 DFTYQKNSVTGFSNKGLOKVENKNLEIPKQHNGVTITEIGNAFRNV-----DFQNKTL 465  
DB 1286 ---TEEKNAKOL-----VLKAKEKAYQDILNAGTNDVTQIKDQAVDQGGTADTTIKDV 1339  
QY 466 RKYDL-----BEVLPSTI-----RKIGAFAPQSN-NLKSPEASDDLEEIKEG 507  
DB 1340 AKDELATKAREQKALIAQTADATTEKEQANQOVDAELTQGNQNIENASIDVNTAKDN 1399  
QY 508 AFM-----NNRIEPL-ELKOKLVITIGDAAFHNIHIAIVLPESVQETG--RS 551  
DB 1400 AIQAIDPIQASTDVKTNARAEILLTEMQNKITEILANNETTNE-----EKGNDIGPVA 1452  
QY 552 APRQNGANNLIPMGSKVKTIGEMAFLSNRLEHLDLSEOKQLTEIPVQAFSONALKEVLLP 611  
DB 1453 AY-REGLANI-----NAAATTG-----DVTAKOTAVQKVQQLHANPVKK---P 1492  
QY 612 ASLATTIREAF--KGNHLKQL-----EVASALSHIAFNALDDNDGDQFQPNK 656  
DB 1493 AG-KTALQQAADKKTQIEQTPNASQOEINDAKQEVDTLQNAQTN-IDQSSTDEYVNA 1550  
QY 657 V-----VVKT--HNSYALADGEHP-----IVDPD-----KLSSTIVDL-EKILKLTG 697  
DB 1551 VKEGKAKINAVKTPSEYKDALAKTEAAYNAKAVTEADNSNASTSEIAEAKQKLAELKQT 1610  
QY 698 LDYSTLROTTQTPQDMTTAGKALLSKNLRGKQKFLQRAQFPFL-----GRVLDLKAIA 753  
DB 1611 ADQNVNQTASKDDIEVQIHNDLNDINDYTIPTGKESATTDUYAYADQKKNISADTNAT 1670  
QY 754 KAEKALVTKATKNGQLLERSINKAV-----LAYNNSAI-----KKAN----- 791  
DB 1671 QDEKQQAIKQVDNQVQTALESINNGVNDGDVDDALTOGKAAIDAIOVDATVKPKANQVID 1730  
QY 792 -----VKLEKELDLTG-----LVEGKGPLAQNTWQGVVLKTPLPPEY 833  
DB 1731 AKAEETKESIDQSDQLTAEEKTEALAMIKQITDQAKQGITDATTTAEVEKAKAQ----- 1784  
QY 834 YIGL-----NVYPDKSGK-----LIYALDM-----SDTIGEGQKDAYGNPILNVBD 875

## RESULT 48

US-10-369-493-22285  
; Sequence 22285, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 22285  
; LENGTH: 1875  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-22285

Query Match 3.9%; Score 196.5; DB 15; Length 1875;  
Best Local Similarity 20.8%; Pred. No. 0.0073;  
Matches 242; Conservative 164; Mismatches 378; Indels 379; Gaps 59;

QY 34 KQTQASSI-SGADYABSSGSKLKINET-----SGPVDDTVDL-----FSDKRT 78  
DB 199 RKTQELTLQSNNDWLEKELRSK---NEQYLSYRQKTDKVIDIRNELNRLNDFQMBT 255  
QY 79 TPEKIK---DNLAGPREQ--ELKAVTENTSEKQITSG-----SOLEQSK 120  
DB 256 NNDVLKQNNLSKSLQKLEIKGLSDLSLSEKQEFSAEMSLKORLVLDLLESQNAVKE 315  
QY 121 SLS-----LN--KTVPSTSNMIECDFITKGNLTVLGSKSGVEKLSQTDHVLPSQA----- 169  
DB 316 ELNSIRELNTAKVIADDSKKQ---TPENEDLLKELQLTKEKLAQCEKELSSITDEA 371  
QY 170 -ADGTQLIQVASFAPDQKTAIAEYTSRAGENGESIQDVGKE---IIN----- 216  
DB 372 DEDNENLSAKSSDFIPLKQLIKERRTKHQLQIETFIVELEHKVPIINSFKERTDML 431  
QY 217 EGEVFNVSLLKQVITPTGYKHIGODAFVNDKNIAEVLNPSLETISDYAFALHALKQIDL 276  
DB 432 ENELNNAALL-----LEHTSNE---KNAKVLELNKQO-----KLVEC 466  
QY 277 PDNLKAIGELAFFNQITGKLSPLRQLMRLAERAFKSNHIK-----TIEF----- 321  
DB 467 ENDLQTLTK-----QRDLRCRQIQYLLITNSVNSDKGPLRKEBEIQIQLIMQEDD 517  
QY 322 ----RGNLSKVGISAFQDNDLSQM-----LPGLE-----KIESE 354  
DB 518 STITESDSQKVYTERLVEFKNIIQLQEKNAELLKVRNLADKLESKEKKSKQSLQKIESE 577  
QY 355 APTGNPGDDHNNRVVLWTSKGNPSGLATENTYVNPDKSLWQSPSEIDYTKWLEEDFTY 414

Db 578 --TVNEAKE-----AIITLKSEK-----MDLESRIEELQELBELKTS 613  
 Qy 415 QKNSVTGPNKGLQKVRNK-NLEIPKOHNGVTITEIGNAFRNVDFQNKTRKYDLLEV 473  
 Db 614 VPNEADASYNVITKQLETYKRDLESQVQDLQTRISQITRESTENSLLNKEIQ--DLYDS 671  
 Qy 474 KLPSTIRKIG-----AFQSNLNKSFASDDLEEKEGAFMNNRI-----514  
 Db 672 KSDISI-KLGKESKRILAEERFKLLSNTLDLTKAEND-QLRKRFPDYLQNTILKQDSKTH 729  
 Qy 515 -----ETLEKDKLVITGDAAFHI--NHIYAVLPESV-QEIGRSAPRQNGANNLIFMGSK 567  
 Db 730 ETLENYVSKSLSVETELLLKKEQKLVHLEKNLQELNKLSPK--DSLRIWVTO 786  
 Qy 568 VKTIGEMAFNLRLHLDLSEQKLTETIPVQAFSDNALKEVLLPASLTITREAEKKNHL 627  
 Db 787 LQTLQKER--EDLLETRKSCQKIDEL-----EDALSE-----LK--KETSQKDHHI 830  
 Qy 628 KOLEVASALSHIAFNALDNDGDEQFQDNKV--VVKTHNSYALADGEHFI VDPDKLSSTI 685  
 Db 831 KOLEE-----DNNSNIEWQNKIEALKKDYESVITSVDSKO--TDIEKLOYKV 876  
 Qy 686 VDLEK-----ILKLEIG-----LDYSTLRQTT---707  
 Db 877 KSLEKEIEEDKIRLHTYNNVMDETINDDSLRKELEKSKINLTDAYSQIKEYKDYETTSQS 936  
 Qy 708 --QT-----QFRDMTTAGKALL--SKSNLRQGEKQKFLQEAOFPL-GRVDLDKATAKAEK 757  
 Db 937 LQOTNSKLDESFKDFTNQIKNLTDEKTSLE--DKISLLKEQMFNLNLDLQKGMEXEK 994  
 Qy 758 ALVTYKATKNGQLLERSINKAVLAYNNSAIK-----KANVKRLEKELDLTLGLVEGKGP 811  
 Db 995 ADFKKRI-----SILQNNKEVEAVKSEYESKLSKI QNDLDOQT-----1033  
 Qy 812 LAQATWQGVYLLKTLPLPEYIYGLNVYFDKSGKLIYALDMSDTIGE--GOKDAYGNPI 869  
 Db 1034 -IYANTAQNNY-----EQELQKHADVSKTISELREQHTYKGV 1071  
 Qy 870 --LVNDED-----NEGYHALAVATLADYEGLDIKTILNSKLQSLTSIRQVPTAAHYR 919  
 Db 1072 KTLNLSRDLQENALKENKESWSQKESL--LEQLDLS--NSRIEDLSQNKL---LYDQ 1123  
 Qy 920 AGIFOAIQNAAAEAEQLPKPGTHS-----EKSSSES 952  
 Db 1124 IQIYTA--ADKEVNSTNGPGLNLIITLRRERDILTKVTVAERDAKMLRQKLSMDV 1180  
 Qy 953 ANSKDRGLQSNPKTRGRHSAIL 975  
 Db 1181 ELQDARTKLDNSRVEKENHSSII 1203

## RESULT 49

US-10-732-923-3334  
 ; Sequence 3334, Application US/10732923  
 ; Publication No. US20050108791A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edgerton, Michael D  
 ; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
 ; FILE REFERENCE: 38-15(52796)C  
 ; CURRENT APPLICATION NUMBER: US/10/732,923  
 ; CURRENT FILING DATE: 2003-12-10  
 ; PRIOR APPLICATION NUMBER: 10/310,154  
 ; PRIOR FILING DATE: 2002-12-04  
 ; NUMBER OF SEQ ID NOS: 24149  
 ; SEQ ID NO 3334  
 ; LENGTH: 1875  
 ; TYPE: PRN  
 ; ORGANISM: Saccharomyces cerevisiae  
 US-10-732-923-3334

Query Match 3.9%; Score 196.5; DB 17; Length 1875;  
 Best Local Similarity 20.8%; Pred. No. 0.0073;

Matches 242; Conservative 164; Mismatches 378; Indels 379; Gaps 59;

Qy 34 KOTQASSSI-SGADYAESSGSKLKINET-----SGPVDVTDL-----FSDKRT 78  
 Db 199 RKTQELTULQSNNDWLEKELRSK---NEQYLSYRQKTDKVIDIRNELNRLNDFQMKRT 255  
 Qy 79 TPEKIK---DNLAKGPREQ--ELKAVTENTSEKQITSG-----SOLEQSKE 120  
 Db 256 NNDVLKQKNNELSKSLQELKLEIKGLSDLSNSEKQEPSAEMSLKQALVDLLESQNAVKE 315  
 Qy 121 SLS-----LN--KTVPSTSNWEICDPITKGNITVLGLSKSGVEKLSQTDHLVLPQA--169  
 Db 316 ELNSIRELNTAKVIADDSKKQ---TPENEDLLKELQLTKEKLAQCEKCECLSSITDEA 371  
 Qy 170 -ADGTQLLIQVASFAPTPDKTAIABYTSRACENGESIQLDVDGKE--IIN-----216  
 Db 372 DEDNENLSAKSSDPIFLKQKLIKERRTKHEHQNOIETFI VELEHKVPIINSFKERTDML 431  
 Qy 217 EGEVFNLYLLKKVTIPTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAFALHALKQIDL 276  
 Db 432 ENELNNAALL-----LEHTSNE--KNAKVKELNAKQ-----KLVEK 466  
 Qy 277 PDNLKAI GELAFFNQITGKLSLPRQLMRLAEAPKSNHIK-----TIEP-----321  
 Db 467 ENDLQTLTK-----QRLLDCRQIQYLLITNSVNSDKGPLRKEBIEQITQIMQEDD 517  
 Qy 322 ---RGNLSKVI GEASFQDNDLSQLM-----LPDGLE-----KISE 354  
 Db 518 STITESDQKVVTERLVEFKNIILQEKNAELLKVVRLADKLESKEKSKOSLOKISE 577  
 Qy 355 AFTGPGGDHYNRNVLTGSKGNPSGLATENTYVNPDKSLWQESPEIDYTKWLEBEDFY 414  
 Db 578 -TVNEAKE-----AIITLKSEK-----MDLESRIEELQELBELKTS 613  
 Qy 415 QKNSVTGPNKGLQKVRNK-NLEIPKOHNGVTITEIGNAFRNVDFQNKTRKYDLLEV 473  
 Db 614 VPNEADASYNVITKQLETYKRDLESQVQDLQTRISQITRESTENSLLNKEIQ--DLYDS 671  
 Qy 474 KLPSTIRKIG-----AFQSNLNKSFASDDLEEKEGAFMNNRI-----514  
 Db 672 KSDISI-KLGKESKRILAEERFKLLSNTLDLTKAEND-QLRKRFPDYLQNTILKQDSKTH 729  
 Qy 515 -----ETLEKDKLVITGDAAFHI--NHIYAVLPESV-QEIGRSAPRQNGANNLIFMGSK 567  
 Db 730 ETLENYVSKSLSVETELLLKKEQKLVHLEKNLQELNKLSPK--DSLRIWVTO 786  
 Qy 568 VKTIGEMAFNLRLHLDLSEQKLTETIPVQAFSDNALKEVLLPASLTITREAEKKNHL 627  
 Db 787 LQTLQKER--EDLLETRKSCQKIDEL-----EDALSE-----LK--KETSQKDHHI 830  
 Qy 628 KOLEVASALSHIAFNALDNDGDEQFQDNKV--VVKTHNSYALADGEHFI VDPDKLSSTI 685  
 Db 831 KOLEE-----DNNSNIEWQNKIEALKKDYESVITSVDSKO--TDIEKLOYKV 876  
 Qy 686 VDLEK-----ILKLEIG-----LDYSTLRQTT---707  
 Db 877 KSLEKEIEEDKIRLHTYNNVMDETINDDSLRKELEKSKINLTDAYSQIKEYKDYETTSQS 936  
 Qy 708 --QT-----QFRDMTTAGKALL--SKSNLRQGEKQKFLQEAOFPL-GRVDLDKATAKAEK 757  
 Db 937 LQOTNSKLDESFKDFTNQIKNLTDEKTSLE--DKISLLKEQMFNLNLDLQKGMEXEK 994  
 Qy 758 ALVTYKATKNGQLLERSINKAVLAYNNSAIK-----KANVKRLEKELDLTLGLVEGKGP 811  
 Db 995 ADFKKRI-----SILQNNKEVEAVKSEYESKLSKI QNDLDOQT-----1033  
 Qy 812 LAQATWQGVYLLKTLPLPEYIYGLNVYFDKSGKLIYALDMSDTIGE--GOKDAYGNPI 869  
 Db 1034 -IYANTAQNNY-----EQELQKHADVSKTISELREQHTYKGV 1071  
 Qy 870 --LVNDED-----NEGYHALAVATLADYEGLDIKTILNSKLQSLTSIRQVPTAAHYR 919  
 Db 1072 KTLNLSRDLQENALKENKESWSQKESL--LEQLDLS--NSRIEDLSQNKL---LYDQ 1123



```
QY 920 AGIFOAIONAARAEABQLPKPQTHS-----EKSSSES 952
Db 1124 IQIYTA---ADKEVNNSNGPGLNLTILRRERDILDTKVTVAAERDAKMLRQISLMDV 1180
QY 953 ANSKDRGLQSNPKTNRGHSAIL 975
Db 1181 ELQDARTKLDNSRVEKENHSSII 1203

RESULT 50
US-10-732-923-3335
; Sequence 3335, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3335
; LENGTH: 1875
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-732-923-3335

Query Match 3.9%; Score 196.5; DB 17; Length 1875;
Best Local Similarity 20.8%; Pred. No. 0.0073;
Matches 242; Conservative 164; Mismatches 378; Indels 379; Gaps 59;

QY 34 KQTQASSI-SGADYAESGSKKINET-----SGPVDVTDVL-----FSDKRT 78
Db 199 RKTQELTLQSNNDWLEKLSK---NEQYLSYRQKTKVILDIRNELNRLRNDQMERT 255
QY 79 TPKEIK---DNLAKPREG---ELKAVTENTSEKQITSG-----SOLEQSK 120
Db 256 NNDVLKQKNELSKLQSKLEIKGLSLNSEKOEFSAEMLSKQRLVDLLESQNLAYKE 315
QY 121 SLS-----LN---KTVPSTGNWEICDFTKGNLTVLGSKGVEKLSQTDHLVLPQSA----- 169
Db 316 ELNSIRELNTAKVIADDSKKQ-----TPNEDLLKELQITKEKLAQCEKECLRLSITDEA 371
QY 170 -ADGTQLIQVASFATPPDKKTAIABYTSRAGENGESIQDVGKE---IIN----- 216
Db 372 DEDNENLSAKSSDFILKQKLIKERRTKEHLQNIETPIVELEHKVPIINSFKERTDML 431
QY 217 EGEVENSILLKVTIPTGKYHGQDAFVDNKNIAEVNLPESLETISDYAPAHLALKQIDL 276
Db 432 ENELNNAALL-----LEHTSNB---KNAKVKELNAKQ-----KLVEC 466
QY 277 PNLKAIGELAFFDQNIQTKLSLPQLMLAERAFKSNHK-----TIEF----- 321
Db 467 ENDLQTLFK-----QRDLCHQIQYLLITNSVNSDKGPKRKEBIOPTIOMQBEDD 517
QY 322 ----RGNSLKVIGESFQNDLSQLM-----LPDGLE-----KIBSE 354
Db 518 STITESDSQKVTERLVEPKNIILQOEKNAELLKVVRNLADKLESKEKSKSQSLQKIBSE 577
QY 355 AFTGNPGDDHNNRVVLTKSGKNSGLATENTYNNPKSLWQESPEIDYTNLEEDFTY 414
Db 578 --TVNEAKE-----AIITLSEK-----MDLESRIEELQELPELKTS 613
QY 415 QKNSVTGFSNKGLOKVRNK-NLEIPKQHNGVTITEIGNAFRNVDFONKTLRKYDLEEV 473
Db 614 VPNEADASVNTIKOLTETKRDLESQVQDLQTRISQITRESTENSLNKSIQ--DLYDS 671
QY 474 KLPSTIRKIG-----APAFQSNLKSFEASDLEBEIKEGAFMNNRI----- 514
Db 672 KSDISI-KLGKESRILAEERFKLLSNTLDLTKAEND-QLRKRPDYQLNTILKQDSKTH 729

QY 515 -----ETLELKDQLVTIGDAAPHI--NHIIAIVLPESV-OEIGRSAPFRQNGANNLIFMGSK 567
Db 730 ETLNEYVCSKSLSISETELLNLKEBQKRLRVHLEKNLQELNKLSPK---DSLRIMVYTQ 786
QY 568 VKTIGEMAFSLNRLHLDLSEQKOLTPIPVQAFSDNALKEVLLPASLKTIRREAFKKNHL 627
Db 787 LQTLQKER--EDLLETRKSCQKIDEL-----EDALSE-----LK--KETSQDHHI 830
QY 628 KQLEVASALSIAFNALDNDGDQFQNKV--VVKTHNSYALADGBHFIVDPDKLSTI 685
Db 831 KQLEE-----DNNSNIEWYQNKIEALKDYESVITSVDSKQ--TDIEKLQYKV 876
QY 686 VDLEK-----ILKLIIEG-----LDYSTLRQTT--- 707
Db 877 KSLEKEIEEDKIRLHTYNNVMDETINDDSLRLKEKSKINLTDAYSQIKYKDLTYETTSQS 936
QY 708 --QT-----QFRDMTTAGKALL--SKSNLRQGEKQKFLQEAQFPL-GRVDLDKAIAKAEK 757
Db 937 LQQTNSKLDSEFKDFTNQIKNLTDEKTSLE--DKISLLKEQMFNLNNELDLQKKGMEKEK 994
QY 758 ALVTKATKQGLLERSINKAVLAYNNSAIK-----KANVKRLEKELDLTLGLVEGKGP 811
Db 995 ADFPKRI-----SILQNNKEVEAVKSEYESKLSKIQNDLDQOT----- 1033
QY 812 LAQATWQGVYLLKTPLPPEYIYGLNVYPDKSGKLIYALDMSDTIGE--GQKDAYGNPI 869
Db 1034 -IYANTAQNNY-----EQLQKHADVSKTISELREQLHTYKQV 1071
QY 870 --LNVDEB-----NEGYHALAVATLADYEGDIDKTLNLSKLSQTSIRQVPTAAVHR 919
Db 1072 KTLNLSRQLENALKENEKSWSSQKESU--LEQLDLS---NSRIEDLSSQNK---LYDQ 1123
QY 920 AGIFOAIONAARAEABQLPKPQTHS-----EKSSSES 952
Db 1124 IQIYTA---ADKEVNNSNGPGLNLTILRRERDILDTKVTVAAERDAKMLRQISLMDV 1180
QY 953 ANSKDRGLQSNPKTNRGHSAIL 975
Db 1181 ELQDARTKLDNSRVEKENHSSII 1203

RESULT 51
US-10-831-070-6
; Sequence 6, Application US/10831070
; Publication No. US20050112612A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Alterman, Eric
; APPLICANT: Cano, Raul J.
; APPLICANT: Hamrick, Alice
; TITLE OF INVENTION: Lactobacillus Acidophilus Nucleic Acid
; TITLE OF INVENTION: Sequences Encoding Cell Surface Homologues and Uses
; TITLE OF INVENTION: Therefore
; FILE REFERENCE: 5051.690
; CURRENT APPLICATION NUMBER: US/10/831,070
; CURRENT FILING DATE: 2004-04-23
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 2539
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-10-831-070-6

Query Match 3.9%; Score 196; DB 17; Length 2539;
Best Local Similarity 19.1%; Pred. No. 0.012;
Matches 237; Conservative 178; Mismatches 440; Indels 388; Gaps 59;

QY 27 LVKPEILKQTOASSISGADYAE-----SSGKSKLKINETSGP-----VDPTVT 70
Db 507 LIKEATDAANNAKEAIDKATTAIDAQDEBGTNNINNVTPVPSLEDAKKAATKAVDADALT 566
```



Db 772 NLELHISKLEHENVSELSEFISGLESQLTLYLANEKELSMLOMDESRLITNLKDE-LEQVE 830  
Qy 432 RNK-NLEIPKQNGVTIIEIGNAFNVDFOKNTLR-----KYDLERVKLPS 477  
Db 831 AQKVELKLOMDESRLITNLKDE-LEQVEAQKVELKLOMDESRLITNLKDELEQVE--- 886  
Qy 478 TIRKIGAFAPQSNLKSFP-----EASDLEEIEKEG-----AFNNRIE-----TLBK 520  
Db 887 -AQY---ELKENQLESHRRLSEVQDSSEALRSNAKLQATVDHVVECKSLQTLTADLK 942  
Qy 521 DKLVITGDAAPHINHIYAIVLPESVQETGRSAFRQNGANNLIFMGSKVKTLGEM-AFLSN 579  
Db 943 KQKLEV-----HGYASHLEQELQSKKTM-----DFCKTLESLEAKLSS 982  
Qy 580 RUEHLDSLQKQLTIP---VQAFSDNAKEVLLPASLKTIREAPKQNLKQL-----EVAS 634  
Db 983 LQEDISLKEQSLSELENIQFQSHKEHERIDRVHLLNKIEKTEVLFHPRFLEREVIS 1042  
Qy 635 ALSHTAFNALDNDQDEQFNKVVVTHHNSVALADGEHFIVDPKLSSTIVDLE----- 689  
Db 1043 LTAQL-----SSTEERBSSTLDTIREVSLRADKAKLEANLEDVNAQMIHYESQLED 1095  
Qy 690 -----KILKLEGL-----DYSTLRQTTQTFQDMTTAGKALLS---KSNLR 728  
Db 1096 LRESKTKIKOLVDSLNASQNEEMLTDDVNNRRIEARSNEDNLRKTLCELELKSXS 1155  
Qy 729 QSEKQKF-----LQEAQFFLGRVD-LDKAIA-----K 754  
Db 1156 DYEQQIITEISVLKIQVHKIAGLQDEVLTQLQSSLEAKFEKGLQGLIQSLSECEBLK 1215  
Qy 755 AKALVTKK-----ATQOQLLERSINKAVLAYNNSAIKANVKELEKELDLLTGL 805  
Db 1216 AQKGLMTDKVSCMQDTLNAANEKGQIEISAQTKVLMGDEPPVKETSDVLEABLKSLSI 1275  
Qy 806 VEGKPLAQATVQGVYLLKTPLPPEYVIGLVNVPFDSKGLIYALDMSDTIGEQKDAY 865  
Db 1276 IRG-----ANSEYQKIYSLQEE---NEDLTRNQLMEKE-----LDUKTS 1313  
Qy 866 GNPILNVBDENGYHALAVATLADYEGLDIKTILNSKLSQLTIRSIRQVPTAAYHRAGIFQA 925  
Db 1314 ----QNKDENTNKQVSLQDEVLMQLSSLDLQSSLEAKFENGKLEGL----- 1350  
Qy 926 IQNAAAEQOLPKPGTHSEK-SSSESANSKDQGLQ 961  
Db 1351 LQSLSECEBELKAQKGLMTDKVSCMQDTLNAANEKGQ 1387

## RESULT 53

US-10-282-122A-60961  
Sequence 60961, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA 034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727

Qy 44 GADYAESGSKKLINETSQVDDVTDLDFDKKTT--PEKIKD-NLAKPREOELKAVT 100  
Db 27 GADITTVSEDTAVKTAADSADEGISETGSDDETAEPKAEKAEKETEKEEKAKT 86  
Qy 101 ENTESEKQI---TSGSQLEQSKESLSLNKTVPTSTSNMEICDPITKGNLTV---GLSKSGV 154  
Db 87 EEPASNKITEINTDKSQLKQI---SLKAAVPAGSTV-----NSLPPDDNLAK--- 130  
Qy 155 EKLQTDHLVPSQAADGTQLIQVASFAFTPKKTAATABYTSRAGEINGEISQLDVDGKEI 214  
Db 131 -KLAV---ITGNAATAGNESVDSAALL-----AISQDL----- 161  
Qy 215 INEGEVFNLSYLLKKVTPPTGYKHIGQDAFVFNKNAEVLNLPESLETTISDYAFAPHLAKQI 274  
Db 162 --SGETND-----PTDISNIEGLQYLE--NLTSNLSEN--NISDLA---PLK-- 201  
Qy 275 DLPNLKAIGELAPFDNQITGKLSLPQLMRLAERAPKSNHIKTIE--FRONSLLKVGEA 332  
Db 202 ---DLVNLVSLNLSNRNLTNLVSGVEDLVNLQELNVSAN--KALEDISQVASLPVLKEI 255  
Qy 333 SPQNDLSQLMLPDGLEKIESEAFGNPGDDHNNRVVLTGSKGNPSGLATENTYVNP 392  
Db 256 SAQGCNKTLEL-----KNPAGAVL----- 275  
Qy 393 KSLWQESPEIDYTKWLEEDPTYQKNSVTGFSNKGLOKVRNKNLEIP-----KOHNG 444  
Db 276 -----PEL-----ETFYQENDLTNLTS--LAKLPKLNLYIKGNASLSKSETLNG 319  
Qy 445 VT-----ITRIGD-----NAPRNV-DFQNKTLRKVDL 470  
Db 320 ATKQLIDASNCTDLETGLDISGLSELEMIQLSGCSKLKEITSLKNLPNLVNITADSCAI 379  
Qy 471 BEVKLPSTIRKIGAFAPQSN-NLKSFEASDLEBK-----EGAFNN-----RIETL 517  
Db 380 EDLGTLANLPKLTQLVLSDNENLTNITAITDLPQKTLTLDGCGITSIGTLDNLPKLEK 439  
Qy 518 ELKDKLVITGDAAPHINHIYAIVLPESVQETGRSAFRQNGANNLIFMG--SKVKTGLGMA 575  
Db 440 DLKE-----NQITSI---SEITDLPRLSYLDVSVNNLTITGDLKCLPLEHLN 484  
Qy 576 FLSNRLHLDLSEKQLTETIPVQAF---SDNALKEVLLPASLKTIREAFKKNHLKQLEV 632  
Db 485 VSSNR-----LSDVSTLTNFPPLNVIINNNVIRTVGQWTELPSELKEFYAQNNSIDISM 539  
Qy 633 ASALSHI-----AFNALDDNDQEQFDNKVVVTHHNSVALADGEHFIVDPKLSSTIV-- 686

Query Match 3.8%; Score 195.5; DB 15; Length 1778;

Best Local Similarity 20.2%; Pred. No. 0.0078;

Matches 209; Conservative 137; Mismatches 330; Indels 361; Gaps 50;

Qy 44 GADYAESGSKKLINETSQVDDVTDLDFDKKTT--PEKIKD-NLAKPREOELKAVT 100

Db 27 GADITTVSEDTAVKTAADSADEGISETGSDDETAEPKAEKAEKETEKEEKAKT 86

Qy 101 ENTESEKQI---TSGSQLEQSKESLSLNKTVPTSTSNMEICDPITKGNLTV---GLSKSGV 154

Db 87 EEPASNKITEINTDKSQLKQI---SLKAAVPAGSTV-----NSLPPDDNLAK--- 130

Qy 155 EKLQTDHLVPSQAADGTQLIQVASFAFTPKKTAATABYTSRAGEINGEISQLDVDGKEI 214

Db 131 -KLAV---ITGNAATAGNESVDSAALL-----AISQDL----- 161

Qy 215 INEGEVFNLSYLLKKVTPPTGYKHIGQDAFVFNKNAEVLNLPESLETTISDYAFAPHLAKQI 274

Db 162 --SGETND-----PTDISNIEGLQYLE--NLTSNLSEN--NISDLA---PLK-- 201

Qy 275 DLPNLKAIGELAPFDNQITGKLSLPQLMRLAERAPKSNHIKTIE--FRONSLLKVGEA 332

Db 202 ---DLVNLVSLNLSNRNLTNLVSGVEDLVNLQELNVSAN--KALEDISQVASLPVLKEI 255

Qy 333 SPQNDLSQLMLPDGLEKIESEAFGNPGDDHNNRVVLTGSKGNPSGLATENTYVNP 392

Db 256 SAQGCNKTLEL-----KNPAGAVL----- 275

Qy 393 KSLWQESPEIDYTKWLEEDPTYQKNSVTGFSNKGLOKVRNKNLEIP-----KOHNG 444

Db 276 -----PEL-----ETFYQENDLTNLTS--LAKLPKLNLYIKGNASLSKSETLNG 319

Qy 445 VT-----ITRIGD-----NAPRNV-DFQNKTLRKVDL 470

Db 320 ATKQLIDASNCTDLETGLDISGLSELEMIQLSGCSKLKEITSLKNLPNLVNITADSCAI 379

Qy 471 BEVKLPSTIRKIGAFAPQSN-NLKSFEASDLEBK-----EGAFNN-----RIETL 517

Db 380 EDLGTLANLPKLTQLVLSDNENLTNITAITDLPQKTLTLDGCGITSIGTLDNLPKLEK 439

Qy 518 ELKDKLVITGDAAPHINHIYAIVLPESVQETGRSAFRQNGANNLIFMG--SKVKTGLGMA 575

Db 440 DLKE-----NQITSI---SEITDLPRLSYLDVSVNNLTITGDLKCLPLEHLN 484

Qy 576 FLSNRLHLDLSEKQLTETIPVQAF---SDNALKEVLLPASLKTIREAFKKNHLKQLEV 632

Db 485 VSSNR-----LSDVSTLTNFPPLNVIINNNVIRTVGQWTELPSELKEFYAQNNSIDISM 539

Qy 633 ASALSHI-----AFNALDDNDQEQFDNKVVVTHHNSVALADGEHFIVDPKLSSTIV-- 686

Db 540 IHMPNLRKVDASNNLTNIG--TFDN-----LPKLSLDVHSNRTSTSVIH 585  
Qy 687 DLKILKLEGLDYSTLROTTQTPRDMTTAGKALLSKNLRQGEKQFLQEAQFPGLGRV 746  
Db 586 DLPSL-----ETFNQTNLTNIG--TMDNLPD-----LTVNLSFNRI 622  
Qy 747 -----DLDAKAIAKALVTKKATNGQLLERSINKAVLAYNSAIKKANVRLE 796  
Db 623 PS LAPIGDLPNLETLLVSDNNLSYLSLGTMDG-----VPKLRILDLQ 664  
Qy 797 KELDLTLGLVEGKPLAQATWQGVYLLKTPPLPYIYGLNVPFK-SG-----KLIYA 850  
Db 665 NNYLNTYGTGNLSSLSLT-----NLTELNRNNVYIDDISGLSLTSLRIY- 711  
Qy 851 LDMSDTIGEGQDAYGNPILNVDEN-EGVHALA-----VATLADYEG 892  
Db 712 -----LNLSNKIEDISALSNTLNQELTLENKKNENISALSDELN 752  
Qy 893 LDIKTLNLSKLSQTSIROVPTAAYHRAGIFQA-----IQNAAAEAEQLLPKP 940  
Db 753 LNKLVVSKNKIIDISPVANN-----VNRGAIVTASNQTYTLPTVLSYQSSFTIDNPVWYD 808  
Qy 941 GTHSEKSSSESANSKD 957  
Db 809 GTLLAPSSIGNSGNKYD 825

RESULT 54  
US-10-724-972A-5942  
; Sequence 5942, Application US/10724972A  
; Publication No. US20040147734A1  
; GENERAL INFORMATION:  
; APPLICANT: Doucette-Stamm, Lynn  
; APPLICANT: Bush, David  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: PATH03-16  
; CURRENT APPLICATION NUMBER: US/10/724,972A  
; PRIOR FILING DATE: 2003-12-01  
; PRIOR APPLICATION NUMBER: 09/450,969  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 7544  
; SEQ ID NO 5942  
; LENGTH: 3696  
; TYPE: PRN  
; ORGANISM: S.epidermidis  
US-10-724-972A-5942

Query Match 3.8%; Score 195; DB 18; Length 3696;  
Best Local Similarity 18.8%; Pred. No. 0.024;  
Matches 220; Conservative 164; Mismatches 478; Indels 310; Gaps 45;

Qy 10 LUTTVSVVTHNQEVFSLVKEPILKQTOASSISGADYAESGKGLKINETSGPVDRTV 69  
Db 1082 ITHTNVNVQKQPQARQALLAKTNEKQSAINSDEGT--IEEKQATQSLNDAKLADEQI 1139  
Qy 70 TDLFSDKR-----TTPEKIKONLAK-----GPRQELKAVTENTSEKQ 108  
Db 1140 TQAASNQNVNALTIGISNISKIQTNTFTKQOARDQVQKQFQKEAELNSTPHATQDEKQ 1199  
Qy 109 ITSGSOLEQSKESLSLNTKVPSTSNWEICDFITKG-----143  
Db 1200 -DALTELTAQKET-ALNDINQATQNVQVDTALTSGIQTQNVNVRKQAEKTTINDIV 1257  
Qy 144 -----NTLVLSKSGVEKLSQTDHLVLP SQ-----AADGTQLIQV 178

Db 1258 OQHQSIOQNDNDATTTEEKEVANNLVNASQNV--ISKIDNATTNNQIDGIVSDGRQIN- 1314  
Qy 179 ASFAFTP-----DKKTAIAEYTSRAGE-----NGEISQLDVGKEII- 215  
Db 1315 ---ALTPDTSIKRNNAKNDIDIKAAADKKIKQIRINDATDEIEIQEANKRKEAKIEAKONIQ 1371  
Qy 216 -----NEGVEFNSYLLKKVTIPTGYKHIQODAFVDNK-----NIAEVLNPESL 258  
Db 1372 RNSTRDQVNEAKTNGINKENITPATTVKSEARQA-VQNKANEQINHIOQNTPDATNEEQ 1430  
Qy 259 ETISDYAFALHAKQIDLPDNLKAGELAPFNQITGKLSPLQML--BLARAFKSNHI 316  
Db 1431 EAINRVS-AELARVQAQINAEHTTQGVKTIKDDAITSLSRINAQVYVEKESARNAIQKAT 1489  
Qy 317 KTIETFRGNSLVIGASQDNDLSQLMLPDGIEKIESEAFNPGDDHNNRVNVLWTKGS 376  
Db 1490 QOTQFINNDNATDEKEVANNLVITATQKSLDNINSLS-----SNNDVENAKVAG 1540  
Qy 377 KN-----PSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKLQKV 430  
Db 1541 INEIANVLPAVAKSKAKKIDQKLAQINQIQ-----THQ-TATTEEKAAIOLA 1590  
Qy 431 KRNKN--LEIPKQNGVTIITEIGDAPRVDFQKTLKRYDLEEVKLPSTTRKGA--- 484  
Db 1591 NQKSNEARTAIQNEHSNNGVAQAKSNGIHEI-----ELVMPDAHKKSDAKOS 1637  
Qy 485 ---FAFOSNNLKSFEASDDEIEIKEGAFMNNRIETLELKKLVITIGDAAFHINHIYAIV 540  
Db 1638 IDNKYNEQSNTINT--TPDATDEEKQKAL-----DKLKTAKONGY--NKVDQQAQ 1682  
Qy 541 LPESVQEIERSAFRQNGANNLIFMGSKVKVTLGEMAFPLSNRLHDLSEKQQLTEIPVQAF 600  
Db 1683 TNQOVSDAKTEAI--DTITNI--QANVAKPSARVELDSKFEDL--KRQINATP--- 1730  
Qy 601 SDNALKEVLLPASLTIREEPKQNHKQLEVASALSHIAFNALDDNDGDEGFDKNNVVK 660  
Db 1731 --NATEEE-----KQDARTQLRNGKRDVKNLIN-----QDRRDNEV--- 1764  
Qy 661 THNSYALADGEGHFIVDPDKLSSTIVDI-----EKILKLEGL 698  
Db 1765 BOHKNIGLOELETIHANPTRKSDALOELQTKFISQTELINNNKDNATNEEKDEAKRLLEIS 1824  
Qy 699 DYSTLRQTQTQ-----PRDMTTAGKALLSKSNLRQGEKQFLQEAQFPGLGR- 745  
Db 1825 KNKTTITNINQATNNQVNDNAKONGNVEIATIIPTATIKDTAKTADKKAEOQVTTINGNN 1884  
Qy 746 --VLDLKAIAK--AEKALVTKKATKNGQLLERSINKA-----VLAYNN-----SAIKKANVR 794  
Db 1885 DATDEKAEARKLVEKAKIEAKSNITNSDTEREVNGAKTNGLEKINNIOPTQTKTNAQ 1944  
Qy 795 -----LEKELDLT-----GLVEGKPLAQATWQGVYLLKTPPL 830  
Db 1945 EINDKAAQSLIQINNTPDATBEEKQAEATNRVNAQAQAIQINNNNAHSTQEVNESKTNSTA 2004  
Qy 831 PEYITGLNVYFDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDNEGVHALAVATLAD- 889  
Db 2005 TIKSVQPNVI-----KPTAINSLTQEAANNQKTLIGNDGNATDDDEKAAKQLVTKLINEQ 2059  
Qy 890 ----YEGLDIKTILNLSKLSQTSIROVPTAAYHRAGIFQAIONAA--AEAEQLLPKPGTH 943  
Db 2060 IQKIHSTQDNQVNDNVKAQAITAKLINANAHKRQDAINILTNLAESKKSDIRAMQDATT 2119  
Qy 944 SEKSSSESANSKORGLQSNPKTNRGRHSAIL 975  
Db 2120 BEKTAIQSID--DTLAQRNNINGANTNALV 2149

RESULT 55  
US-09-815-242-11501  
; Sequence 11501, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert

```

; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11501
; LENGTH: 949
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11501

Query Match      3.8%; Score 194; DB 9; Length 949;
Best Local Similarity 19.6%; Pred. No. 0.004;
Matches 202; Conservative 150; Mismatches 343; Indels 338; Gaps 48;

QY      6 KTVALTTLTVSVVTHNQ--EVPFLVKEPILKQTOASSISGADYABSSGSKLKLINETSQ 63
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      29 KDIGLELKNPMTPEQAGLYKYIVDIGIKEIQAN-----OQTKN 69
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      64 PVDRTVTLPSDKRTTPEKIDKNLAKPREQELKAVTNTES-EKQITSGSQLESKSL 122
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      70 PQQNKDOL--NTAATPKLAKKASKTPKKEAKQPKKTKKEKKEAPAPIIKKEIE 127
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      123 SLNKTVPSTSNWEICDFTTKGNTLVGLSGVGEKLSQTDHLVLPSSQAADGTQLIQVASFA 182
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      128 IVNTFENQTPLE-----NTPKAVSHSQIEKAKQKLQEIQSREA---LNKLTQSN 175
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      183 FTPDKKTAIAEYTSAGNGEISQLDVPGKEIINEGEVFNYSLLKKVVTIPTGYKHIGQDA 242
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      176 TTTNNANASNVSNNAK--BISEVKKQEQEIKRH-----ENIKRRTGFRVIRKND 224
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      243 FVDKNIAEVLNLPESLETISDYAFAPHALKQIDLDPNLKALGELAFFDNOITGKLSLPRQ 302
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      225 --ETENETENSVESTSKPTQSAALFEDIKK-----EWQEKQETKTKTKRPSK 271
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      303 LMRLAERAFKSNHIKTI-----EPRGNSLKVIGEASFQDNDLSQML-----PDGLEKI 351
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      272 --PRATPTAKNNKSHKIDPSDVDPFGNDI-----YDDETOBILLPDLHEQDNLNKE 321
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      352 ESEAPTPNGDDHYNRVVLWTKSGKNSGLATENTYVNPDKLSQBSPEIDYTKWLBE- 410
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      322 BEE-----KEARQNIINDRVV---QRKNP-----WNEA 347
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      411 DFTYQKNSVTGFSNKGLOKVRKNLEIPKQHGVTITEIGDAPRVV----- 458
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      348 GIKQSKKKKVRFRNDSQKQSA-IAIPEE---VRVYFAQKAMNLNADVIKTLFNLGL 403
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      459 -----DFQNKTLRKDYDLBEVKLPSTIRKIGAFQSNLKLKSFASDDLEEIKE----- 506
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      404 MVTKNDFLKDSEIILAEFFHLEISVQ-----NTLEEFVEEVLGVKKRPPVVT 454
DB      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

; APPLICANT: Hasselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12913
; LENGTH: 949
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-12913

; Sequence 12913, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Hasselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12913

```

```
; LENGTH: 2186
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12913

Query Match      3.8%; Score 194; DB 9; Length 2186;
Best Local Similarity 19.1%; Pred. No. 0.013;
Matches 216; Conservative 183; Mismatches 387; Indels 342; Gaps 58;

QY 20 HNOEVFSLVKEPILKQTAASSISGADYAESSGSKLKNETSGPVDVTDVTLFSD--- 75
DB 1208 YNAKLAIEINATPDATNDEKNAINTLNQDRQQAIESIKQANTNAEVDQAATVAENNIDAV 1267

QY 76 -----KRTTPEKIKONLAKGPREQELKAVTENTESEKQITSGSQLEBSKE-----SLSLN 125
DB 1268 QVDVVKQQAARDKITAEVAK--RIEAVKQTPNATDEEKQ--AAVNQINQLKQDQAINQINON 1324

QY 126 KT---VPSTSNWEI-----CDFITKNTLVGLSGVKEKLSQTHLVLPSSQAADGTQL 175
DB 1325 QTNQDQVDTTNOAVNAIDNVEAEVVIKPKAIDIEKAVKEKQOQIDNSL---DSTDNEK- 1380

QY 176 IQVASFAFTPDKKTATAYTSRAGENGESQLDVGKEIINEGEVFNLSVLLKKVT--IPT 233
DB 1381 -EVASQALAKEKEKALAA-IDQAQTSQVNOQAATNGVSAIK-----IIQPETKVKPA 1430

QY 234 GYKHIGQDAPFVNKNIAEAVNLPE-----SLETISDYAFALHAL-----KQID 275
DB 1431 AREKINQKA--NELRAKINQDKEATEERQVALDKINE--FVNQAMTDTITNRTNQOVD 1485

QY 276 LPDNLKAIGELAFDNOITGKLSLPQMLRAER-----AFKSNHIKTIIEPRGNSLKV 328
DB 1486 -DTSQALDSIA-----LVTDPDHVRAAARDVAKQOYEAKGRIEQAHADEBEQ 1535

QY 329 IGEASFQDND-----LSQLMLPDGLEKIESEAFNGPDGDDHNNRVLVLTGSKNPSGL 382
DB 1536 VALNQLANNEKRALQNDQAIANNQVKTETNGIATLKG---VQPHIVI-----KPEAQ 1587

QY 383 ATENTYVNPDKSLWQESP-----EIDYTKWL-----BEDFTYQKNSVTGFSN--- 424
DB 1588 AIKASAENQVESI-KOTPHATVDELDEANQLISDTLKAQOEIENTNQDAAVTDVRNQT 1646

QY 425 KGLQKVK-----RNKN-----LEIPKOHNGVTTIEGD--NAFRNVD 459
DB 1647 KATEQIKPKVRRKRAALDSIENNNQNOLDAIRNTLDTTQDERVDAIDTLNKIVNTIKNDI 1706

QY 460 FQNTLRKYDLEE-----VKL-----PSTIRKIGAFQFOSNNLKSFEASD-DLEEI 504
DB 1707 AQKNTNAEVRTETDGNNDIKVILPKQVVKPAQSVGVKAEQAQNAL--IDQSDLSSTEE 1764

QY 505 KEGD-----FMNRIETLEBKDLVTIGDAAAFHINHYAIVLP-----ESVQEI 549
DB 1765 RLAAKHLVEQALNQADQINHADKTAQVNOQDSINAQNIISKIIPATTVKATALQOIQNIA 1824

QY 550 RSAPFRONGANN-----LIFMGSKVK---TLGEMAPL-----SNRLEHLD 585
DB 1825 TNKINLIKANNEATDEBQNIATAQVEKELIKAKQOIASAVTNADVAVLLHDEKNEIREIE 1884

QY 586 -----LSEOKQLETPVQAFSNAKVELLPASLKTIREEAPKXHLKQLEVASALS 639
DB 1885 PVINRKASAREQUTTL-----FND---KKQAIERANIQATVEE--RNSILAQLQ---NIYDT 1932

QY 640 AFNALDNDGDQEDFN--KVVVKTHTNSYALADGEHFI VDPDKLSSTIVDLEKILKILIEG 697
DB 1933 AIGQIDODRSNAQVDKTASLNQITIDL-----DVHPKIPDAEKTINDDLARVATLVQ- 1986

QY 698 LOYSTLRQTTQTOFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEK 757
DB 1987 -----NVRKVSNRNK-----ADALK 2001

QY 758 ALVTKATKNGQLLERSIN---KAVLAYNNSATKANVKELEKELDLTLGLVEKGFLAQ 814
DB 2002 AITALKQMDDEELKTARTNADVDVAVLKFNFVALSDIEAVITEKENSLLR--IDN---IAQ 2056
```



```
Db      1707 AONKNAEVDRETGDNKIKVILPKVQVPAAROSVGKAEQAQNAL--IDQSDLSTEEB 1764
QY      505 KEGA-----FNNRIETLEKDKLVTIGDAAFHNIHAIYVLP-----ESVQETG 549
Db      1765 RLAAKHLVQALNQALDQINHADKTAQVQNDQSINAQNIISKIKPATTVKATALQIQINIA 1824
QY      550 RAFAFRONGANN-----LIFMGSKVK---TIGEMAFI-----SNRLEHLD 585
Db      1825 TNKINLIKANNEATDEQNIATAQVEKELIKAKQIQIASAVTNWADVAYLLHDEKNEIRETE 1884
QY      586 -----LSQKQOLTEIPVQAFSDNALKEVLLPASLKTIREEAFFKNHKLQLEVASALSHI 639
Db      1885 PVINRKASAREQUTLL-----FND---KKQAIENIQATVEE---RNSILAQLQ---NIYDT 1932
QY      640 AFNALDDNDGDQFON---KVVVKTTHNSYALADGHEFIVDPDKLSSTIVDLSEKILKLIBG 697
Db      1933 AIGQIDQDRSNAQVDKTAASLNLTIDHL-----DVHPIKKPDAEKTINDDLARVVALVQ- 1986
QY      698 LOYSTLRQTTQTFQFRDMTTAGKALLSKSNLRQEKQKFLQEAQFFLGRVVDLDKAIKAKBK 757
Db      1987 -----NYRKVSNRNK-----ADALK 2001
QY      758 ALVTKKATKNGOLLERSIN---KAVLAYNNSAIKKANVKRLEKELDLTLGLVEGKPLAQ 814
Db      2002 AITALKQDEELKARTWADVAVLKRFNVALSDIEAIVTEKNSLLR--IDN---IAQ 2056
QY      815 ATMVQGVYLLKPLPLPEYIYIGLVYFVFGSKLIYALDMSDTIGEQCKDAYGNPILNVD 874
Db      2057 QT-----YAKFRAIATPEQLAKVKVLIDQ-----YVAD-----GNRM--IDE 2091
QY      875 DNEGHYHALAVATLADYEGDIDIKILMSKLSQTSIROVETAAYHRAGIFQATONAAAEAB 934
Db      2092 D-----ATLN-----DIKQHTQFIVDEILAIAK-LPAEA-----TKVSPKEI 2126
QY      935 QLLPKPQTHSEKSSSESANSKDRGLQSNPKTNRGHSAILLRPTGSKG 982
Db      2127 QPAPKVCTPIKKEETHESKRKE-----LPTGSEG 2158
```

## RESULT 58

```
US-09-815-242-5835
; Sequence 5835, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Tráwick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5835
; LENGTH: 2434
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5835

Query Match      3.8%; Score 194; DB 9; Length 2434;
Best Local Similarity 19.5%; Pred. No. 0.015;
Matches 226; Conservative 177; Mismatches 454; Indels 300; Gaps 52;

QY      1 MKKHLKTLVTLTTVVTVTHQEVSLVKEPIILKQTQASSSI--SCADYAESGK----- 53
Db      226 LKEQVQS--ATTLDGVQTVVNSSTLTNTAMKGLRDSIANEATIKACQNTYDASPNRNEYD 284
QY      54 -----SKLINETSQPV--DDTVTDLFSDKRTTPEKIKDNLAGPRE--QELKAVTENTES 105
Db      285 SAVTAAKAIINQTSNPTPEPNTITQVTSQVTT-----KEQALNGARNLAQAQTTAKNNLN 339
QY      106 EKQITSGSQLEQSKESLSLNTKTVPSTSNWEICDPITKGTNLVGLSKSGVEKLSQTDHL-- 163
Db      340 NLTSINNAQKDALTRSIDGATTVAGV--NQETAKATELNNAHSL--QNGINDETQTKQTK 397
QY      164 ---VLPSQAADGTQLIQVASPAFTDK-----KTAIAEVTSRAGENGISQLDVGKEII 215
Db      398 YLDABEPSKKSAYDQAVNNAKAILTKASQNVDKAAV-----EQALQNVNSTKTALNGDAKL 453
QY      216 NEGEVFNLYLLKKVTIPTGYKHI--GQDAFVNDKNIAEVLNPESLETISDYAPAHALAKQ 273
Db      454 NEAKAAAKQTLGLTL-----HINNAQRTALDNE--ITQATNVGVNTVKAKA-----QQ 500
QY      274 IDLPNLKAIGELAPFDNQITGK--LSLPRQLMRLAERAFKSNHIKTIIEPR-----G 323
Db      501 LD-----GAMGQL-----ETSIRDKDTTLOSQNYQDADAKRTAYSQAVNAAATILNKTAGG 552
QY      324 NSLVKIGASFO-----DNDLSQLMLPDGLEKIBSEAFTEGNDPDHNNRVVL---WTKS 375
Db      553 NTPKADVERAMQAVTQANTALNGIQNLDRAKQAANTAIT--NASDLNTYKQKALKAQVTSQA 611
QY      376 GKNPISGLATENT-----YVNPDKSLWQESPEIDYTKWLE 409
Db      612 GRVSAANGVEHTATLNTAMTALKRAIADAKBTASGNYVNADANKRQA-----YD 662
QY      410 EDFTYQKNSVTG-----PSNKGLO-----KVRKNKULEIPKQH-----NGVT-- 446
Db      663 EKVTAENIVSGTPTFTLTPADVNTAAQTQVNTNAKTQLNGNHNHLEKAKONANTAIIDGLTSL 722
QY      447 -----ITEIGNAFRNVDFQNKTLRYDLEEVKLPS----- 477
Db      723 NGPQKAKLKEQVGQAATLTPNVQTVTRDNA--QTUNTKMGLRDSIANEATIKACQNTYDASQ 781
QY      478 -----TIRK--IGAFAFQSNLNKLSFEASDDLEIEIKGAF---MNNRIETLELKDK 522
Db      782 NKQTDYNSAVTAAKAIIGQTTSPSNAQEIQAQDVTKAQQALNGQENLRPTAQTNAKQH 841
QY      523 LVTIGD-----AAPHNHIYAIVLPSVQEIQRSAFRQNGANNLIIPMGSKV 568
Db      842 LNGLSDLTDAQDKDAVKRQIEGATHVNEV-----TOAQNNADALANTAMTNL 886
QY      569 KT--LGEMAFSLNRLEHLDLSEOKQLTEIPVQAFSDNALKEVLLPASLKTIREEAPK- 623
Db      887 KNGIQDQNTIKQGVNFTDADEAKRNAYTNNAVTOASQILNKAGPNTSKDGVETALENVOR 946
QY      624 --KNHLKQLEVASALSASHIAPNALD-----DNDGDEQFDNKKVVVKTTHNSYALADGHEFIVD 677
Db      947 AKNELNGQNVANAKTTAKNALNLTLSINNAQKALKSQI-----EGATTVAG 994
QY      678 PDKLSSTIVDLEKIL--KLIEGLD-----YSTLRQTTQTFQFRDMTTAGKALLSK-- 724
Db      995 VNOQVSTTASELNTAMSNLQNGINDEAAATKAAQKYTDADREKQTYNDAYVTAATKLLDKTA 1054
QY      725 --SNLRQGEKQKFLQEAQFFLGRVVDLDKAIKAKKALVTKKATKNGOLLERSINKAVLAYN 783
```

Db 1055 GSDNKAQVQALQ-----RVNTAKTALNGDERLNEAKNTAKQV-----ATMSHL 1100  
QY 784 NSAIKXANVRLKELDLTLGLVEGKPLAQ--TMVQGVVLLKTPPLPEYIIGLVNYPD 842  
Db 1101 TDA-QKAN---LTSQIESGTTVAGVQIQANAGTLDQAMNQLRQSTASKDATKSSDYQD 1156  
QY 843 KSKLIYALDMSOTIGEGQKDAYGNPILNVDENE-----GYHALAVA----- 885  
Db 1157 ANADLQWAYNDVTAEGAIIISATNPNMPPDTINQKASQVNSAKSALNGDEKLAQAQTA 1216  
QY 886 ---LADYEGDLIKTLNKSLSQTSIRQVPTAAYHRAGIFQATONAAAEQILPKP 940  
Db 1217 KSDIGRLTLNNAQ--RTAANAQVDOAPNTAAV--TAAKNKA---TSLTAMGNLKHAL--- 1268  
QY 941 GTHSEKSSSESANSKD 957  
Db 1269 ---AEKDNTKRSVNYTD 1282

RESULT 59  
US-10-732-923-3302  
; Sequence 3302, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 3302  
; LENGTH: 1175  
; TYPE: PRT  
; ORGANISM: Clostridium perfringens str. 13  
US-10-732-923-3302

Query Match 3.8%; Score 193; DB 17; Length 1175;  
Best Local Similarity 20.6%; Pred. No. 0.0062;  
Matches 204; Conservative 159; Mismatches 344; Indels 286; Gaps 50;

QY 29 KEPILKQTOASSISGADYAESGSKLKNINETSQVDDTVTLDFSDKRTTPEKIXDNLA 88  
Db 220 KEBEIKEI---NLSIKSEELNLIKKEPEEAE-KVWNTQKELY-DKRIEESLVS--- 270  
QY 89 KGPBRELKAVTENTE-----SEKQITSGSOLEQ-----SKESL---SLNKTVPSTSNWEI 136  
Db 271 ---RSEIKSFKEVBEISNKAQKVIIVINNLEBILKEINKEDLKFSELNKLBEELIN--- 324  
QY 137 CDFITKGNLTL--VGLSKSGVEKL-----SQTDLHLVLPQOAGDTQLIQVAFSA 182  
Db 325 ---LREENKLKPFBEFKKEEKLPLDLRLKKEKLESQKRDILFQTKADGVKLCKACKKI 381  
QY 183 FTPDKK-----TAAEYTSRAGENSISQLDVQKEI-----INEG----- 218  
Db 382 FEDRSKCDTKLNSIENEKELNE--ELKEKEERKEBELFVHEEFKNKINSGLFILNSYESL 439  
QY 219 -EVFNYSLLKKVTIPTGYKHIGQDAFVNDKNI----- 249  
Db 440 DKQFNEIKSEBELKVKYIKNLTEDEKSEKDLKVKVESLKIRDKLESILKETPGDSNGI 499  
QY 250 --NEVNLPSELTISDYAFALHALKQI--DLPNLKAIG-----ELAFFNOITGKLSLPR 301  
Db 500 LEKQIKLGEYREKLNKY-----KEIKNSLESLSKTQNNFEKLTFFNQ---KLLLEK 549  
QY 302 QMLRAERAPKSNHKTIEF-----RGNLSKLVIGEASFODNLSOLMLPDGLEK-- 350  
Db 550 EVRELKDYI---NKVKVEELAHKLRENVLVEGECPCVCGSTHHELNKVEKLNLESEKTI 606  
QY 351 -IESEAFNGPDGDDHNNRVVLTGKGNPKSPGLATENTYVNPDKSLWQESPEI--DYTKW 407  
Db 607 LLESK-----EKLKELIL--EFSKIBATLEYENKKIEELNIEIEEVEGVENERLKF 656

QY 408 LBEEDFTYQKNSVTGFS-----NKGLQKVKRNK-NLE-----IPKOHNGVT 446  
Db 657 LBEENYTLKDKIEFNKLNKKEKLEKOLKLEKKEKNLNLENFNKAEVILCEKIVREKIASK 716  
QY 447 ITEIGD-----NAPRN-----VDFONTKRLKYDLBEVKLPSTIRKIGAFAFOSNN 491  
Db 717 IKELOKELKNSSELNSIKNELKIEDIKLENEBILKKEKKNLLEKEIRILRQTLESNK 776  
QY 492 LKS--PEASDLEE--IKEGAFMNNRIETLEKDKLVT-----IGDAAFHINHIIYALVP 542  
Db 777 INDELREKRDALKENYLSOKSLDGKVEYREKRMKIGSLKGLIDEALPIEKIDIKGLL 836  
QY 543 ESQVEIGRSAPRQNGANNLIFMGSKVKTILGEMAFLSNRLEHLDLSQKQLTETPVOAFSD 602  
Db 837 EDLQL-----EIDYIEK--AYLNLSEAKELE---KAF-N 865  
QY 603 NALKEVLLPA-----SLKTIREEAFKQNLHLKQLEVASALSHIAFNA-LDNDG----- 649  
Db 866 NMNQEVSVTKERVNSLKRKENEKKNL-----ALBEEGFKTILEVKEGILSKDEKE 918  
QY 650 -----DEQFDNKVV-VKTH-----HNSVALADGHEFIVDPDKLSSTIVDLEKILKLE 696  
Db 919 KUKTLIEYDNNLIKIKANIELLIKKLNGKSUTEETWTRVLOEKNN-----EKEUKVE 973  
QY 697 GL-----DYSTLRQTTQTPFRDMTTAGK-----ALLSK-SNLROGEK-QKFLQEAQFFL 743  
Db 974 ELKIRLVTESESIKKLBEQORDILHIKAKQEHKALLSDEKLFGKKFVEFIAANQLKY 1033  
QY 744 GRVDLDKATAKAELAVTKKATNGQLLERSINKAVLAYNNSAIKKAANYKRLKEKLDLLT 803  
Db 1034 ISIEASKKLKDTNGVYGVLEVDENGKFIIRD-----YKNGG-----AERDASTLS 1078  
QY 804 GLVEGKGLAQATMVGQVYLLKTPPLPEYI 835  
Db 1079 GGFTFLASLALSLSSQQLKGTAPLELFL 1110

RESULT 60  
US-10-723-860-862  
; Sequence 862, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Gineburg, Wendy M.  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
; FILE REFERENCE: 05882.0193.NPUS01  
; CURRENT APPLICATION NUMBER: US/10/723,860  
; CURRENT FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: 60/429,739  
; PRIOR FILING DATE: 2002-11-26  
; NUMBER OF SEQ ID NOS: 8393  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 862  
; LENGTH: 3113  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-723-860-862

Query Match 3.8%; Score 192.5; DB 16; Length 3113;  
Best Local Similarity 20.2%; Pred. No. 0.026;  
Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;

QY 6 KTVALTTLTVSVVTHNQEVFLSVKEPILKQTOASSISGADYAESGSKLKNINETSQV 65  
Db 1972 KTTALDQLEKMKKEKQEL-----ESHQSECLHCIOVAEAEVKEKTELLQI---L 2018  
QY 66 DDTVDLFSDKRTTPEKI-----KDNIAKGPBRELKAVTENTESEKQITSGSOLESKES 121  
Db 2019 SSDVSELLKDKTHLQBLEKQSLSKQALSLTKCELENQIALNKKKELL-----VKES 2071

```
QY 122 LSLNKTVPSTSNWEICD-----FITKGNLTVLGLS-----KSGVEKLS---OTD 161
Db 2072 ESIQARL--SESDYEKLVNSKALEALVEGEFALRLSSQBEVHQLRGIEKLRVRIAD 2130
QY 162 -----HLV--LPSQAADGTQLIQVASFAPTPDKKTAIABYTSRAGENGHSIQLDVDG--- 211
Db 2131 EKKQLHIAEKLKERENDSL-----KDKVENLERELQWSEBQELVILDAENSKA 2181
QY 212 -----KEIINEGEVFNLSVLLKKVTIPTGYKHI-----GQ-----DAF 243
Db 2182 EVETLTKQIEEMARSLKVFE---LDLVLRSEKENLTQIQBKQOGLSELDKLLSPFKSL 2238
QY 244 VDNKNAEAVNLPESLETISDYAFAPHLAKQIDLPNLKAIAGELAFPDNQITGKLSLPROL 303
Db 2239 LEEKEQAEIQIEESKTAVEM-----LQNLKEL-----NEAVALCGDQEI 2280
QY 304 MLEAERAFKSNHKTIEFRGNSLKVIGEASFQNDLSQMLPDGLEKIESEAFNGPDD 363
Db 2281 MKATEQSLDPPIEBEHQLR-NSIEKL-RARLEADEKKQLCVLQQLKESEHHA-----D 2331
QY 364 HYNRV-----VLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKN 417
Db 2332 LLKGRVENLERELEIARTNQEHAALEAENS-----KGEVETLKAKIEGTMQ--- 2377
QY 418 SVTGFSNKGLO-----KVKRNKNLEIPKQHNGVTITTEIGDNAPRNVDQFNKTLRKYDL 470
Db 2378 -----SLRGLDVTIRSEKENLTNELOKQERISELEIINSFENI--LQEKQEKVQM 2431
QY 471 BEVKLPSTIRKIGAFAPQSNLKSFEASDDLEBEIK-EGAFMNNRIETLEKDLKVTIGDA 529
Db 2432 KEKS--STAMEMLTQOLKELNERNVAALHNDQEAQKAKQNLSSQVECLEKAQLQGLD 2489
QY 530 AFHINHIYAVLPESV-----QBIGRSAPRQNGANNLIFMGSKVKTLGEMAFUSNLEHLD 585
Db 2490 EAKNNY---IVLQSSVNGLIQEV-----EDGQKL-----EKKDEISRLKNOIQ--- 2531
QY 586 LSEQQLTEIPQAFSDNAL-KEVLLPASLTKTIREAPKKNHLKQLEVASALSHPAFNAL 644
Db 2532 --DQQLVSKLSQVEGEHQLWKEQN--ELRNLTVLEQKIQVLQSKNAS-----L 2578
QY 645 DDNDGDEQDNKVVVVKTHNSYALADGEHFIVDPDKLSSTIVDLBKILKLIBGLDYSTLR 704
Db 2579 QD-----TLEVLQSSYKNLENELELTQMDKMSF---VEKVNKM-----TAK 2616
QY 705 QT-TQTQFDMWTAGKALLSKNLKQGEKQFLQEAQFFLGRVDLDKAIAKAKALVTYK 763
Db 2617 ETELQREHEM--AOKTABELQEL-SGKRNLAGELQLLLEBEI-----K 2657
QY 764 ATKNGOLLERSINKAVLAYNNSAIKKA---NVKRLKELDLTLTGIVRGKGPLAQTWVQ 819
Db 2658 SSKD-OLKE-----LTLENSLKKSLDCMHKQOQVEKEGRVREIAYQLRLHAEKXH 2709
QY 820 GYVLLKPLPLPEYVIGLVNVPDK-----SGKLIYALDMSDTIG 858
Db 2710 QALLDNTN---KQYEVETIYREKLTKEECULSSQKLEIDLKLSKEELNLSKATTQIL 2766
QY 859 EQQKDA-----YGNPILNVDEBNEGYHALAVATLADYEGLDIKTILNSKLSQ----- 906
Db 2767 BELKTKMDNLYVNLKKNENRAGCKMKLLIKSCCKQLE--BEKEILQKELSQLQAQEK 2824
QY 907 -----TSIRQVPTAAVHRAGIQAIONAAAEAEQLLPKG-----THSEKSSSESAN 954
Db 2825 QKTGTVMDTKVDELITTEIKE-----LKETUEEKTKEADEYLDKYCSLLISHLEKAKEMLE 2881
QY 955 SKDRGLQSNPKNTRGRHSAIL 975
Db 2882 TQVAHLCSQSQSDSRGSPLL 2902
```

RESULT 61

US-10-756-149-4927

; Sequence 4927, Application US/10756149

; Publication No. US20050181375A1

```
; GENERAL INFORMATION:
; APPLICANT: Aziz, Nataasha
; APPLICANT: Zlocnik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756,149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 4927
; LENGTH: 3113
; TYPE: PRT
; ORGANISM: Homo Sapiens
; US-10-756-149-4927

Query Match 3.8%; Score 192.5; DB 18; Length 3113;
Best Local Similarity 20.2%; Pred. No. 0.026;
Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;

QY 6 KTVALTTLTVSVVTHNQVFSLVKEPILKQTOQASSISGADYASSSGSKUKINETSQPV 65
Db 1972 KTTALDQLSEKKEKTQEL-----ESHQSECLHCIOVAEAEVKEKTELLQT---L 2018
QY 66 DDTVTDLPFSDKRTTPEKI---KONLAKGPREQELKAVTENTESEKQITSGSQLESQSKES 121
Db 2019 SSDVSELKDKTHLQELQSLSEKDSQALSLLTKCELENQIAQLINKKEKELL-----VKES 2071
QY 122 LSLNKTVPSTSNWEICD-----FITKGNLTVLGLS-----KSGVEKLS---OTD 161
Db 2072 ESIQARL--SESDYEKLVNSKALEALVEGEFALRLSSQBEVHQLRGIEKLRVRIAD 2130
QY 162 -----HLV--LPSQAADGTQLIQVASFAPTPDKKTAIABYTSRAGENGHSIQLDVDG--- 211
Db 2131 EKKQLHIAEKLKERENDSL-----KDKVENLERELQWSEBQELVILDAENSKA 2181
QY 212 -----KEIINEGEVFNLSVLLKKVTIPTGYKHI-----GQ-----DAF 243
Db 2182 EVETLTKQIEEMARSLKVFE---LDLVLRSEKENLTQIQBKQOGLSELDKLLSPFKSL 2238
QY 244 VDNKNAEAVNLPESLETISDYAFAPHLAKQIDLPNLKAIAGELAFPDNQITGKLSLPROL 303
Db 2239 LEEKEQAEIQIEESKTAVEM-----LQNLKEL-----NEAVALCGDQEI 2280
QY 304 MLEAERAFKSNHKTIEFRGNSLKVIGEASFQNDLSQMLPDGLEKIESEAFNGPDD 363
Db 2281 MKATEQSLDPPIEBEHQLR-NSIEKL-RARLEADEKKQLCVLQQLKESEHHA-----D 2331
QY 364 HYNRV-----VLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKN 417
Db 2332 LLKGRVENLERELEIARTNQEHAALEAENS-----KGEVETLKAKIEGTMQ--- 2377
QY 418 SVTGFSNKGLO-----KVKRNKNLEIPKQHNGVTITTEIGDNAPRNVDQFNKTLRKYDL 470
Db 2378 -----SLRGLDVTIRSEKENLTNELOKQERISELEIINSFENI--LQEKQEKVQM 2431
QY 471 BEVKLPSTIRKIGAFAPQSNLKSFEASDDLEBEIK-EGAFMNNRIETLEKDLKVTIGDA 529
Db 2432 KEKS--STAMEMLTQOLKELNERNVAALHNDQEAQKAKQNLSSQVECLEKAQLQGLD 2489
QY 530 AFHINHIYAVLPESV-----QBIGRSAPRQNGANNLIFMGSKVKTLGEMAFUSNLEHLD 585
Db 2490 EAKNNY---IVLQSSVNGLIQEV-----EDGQKL-----EKKDEISRLKNOIQ--- 2531
QY 586 LSEQQLTEIPQAFSDNAL-KEVLLPASLTKTIREAPKKNHLKQLEVASALSHPAFNAL 644
Db 2532 --DQQLVSKLSQVEGEHQLWKEQN--ELRNLTVLEQKIQVLQSKNAS-----L 2578
QY 645 DDNDGDEQDNKVVVVKTHNSYALADGEHFIVDPDKLSSTIVDLBKILKLIBGLDYSTLR 704
Db 2579 QD-----TLEVLQSSYKNLENELELTQMDKMSF---VEKVNKM-----TAK 2616
QY 705 QT-TQTQFDMWTAGKALLSKNLKQGEKQFLQEAQFFLGRVDLDKAIAKAKALVTYK 763
```

Db 2617 ETELOREHEM--AQKTAELOEL--SGEKRLAGELQLLLEI--K 2657  
Qy 764 ATKNGOLLERSINKAVLAYNNSAIKKA-----NVKRLKELDLTLGLVEKGPLAQATWVQ 819  
Db 2658 SSKD-QLKE-----LTLENSELKSLDOMHKDQVEKGVREETAETAYQLRHEAKKH 2709  
Qy 820 GYVLLKTPLPPEYIYGLNVYFDK-----SGKLIYALDMSDTIG 858  
Db 2710 QALLLDTN--KQYEVEIQTYREKLTSSKECLSSQKLEIDLKSSKEELNNSIKATTQIL 2766  
Qy 859 EGOKDA-----YGNPILNVDNEGYHALAVATLADYEGLDIKTILNSKLSQ 906  
Db 2767 BELKTKMDNLKYVNLKKNENARAQGMKLLIKSKCKOLE--BEKETLOKELSQLOAAQEK 2824  
Qy 907 -----TSIRQVPTAAYHRAGIFQAIQNAAAAEQQLPKPG-----THSEKSSSESAN 954  
Db 2825 QKTGTWMDTKVDELTTTEIKE---LKETLEKTKEADEYLDKYCSLLISHEKLEKAKEMLE 2881  
Qy 955 SKDRGLQSNPKTNRGRHSAIL 975  
Db 2882 TQVAHLCSQSQKQDSRGSPL 2902  
  
RESULT 62  
US-10-751-736-99  
; Sequence 99, Application US/10751736  
; Publication No. US20040265230A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Martinez, Robert  
; APPLICANT: Brown, Eugene  
; APPLICANT: Liu, Wei  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING AND TREATING COLON  
; FILE REFERENCE: AM100927 (031896-002000)  
; CURRENT APPLICATION NUMBER: US/10751,736  
; CURRENT FILING DATE: 2003-01-06  
; PRIOR APPLICATION NUMBER: US Provisional Application 60/438,000  
; PRIOR FILING DATE: 2003-01-06  
; NUMBER OF SEQ ID NOS: 54873  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 99  
; LENGTH: 3114  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-751-736-99  
  
Query Match 3.8%; Score 192.5; DB 16; Length 3114;  
Best Local Similarity 20.2%; Pred. No. 0.026;  
Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;  
  
Qy 6 KTVALTTLTVTVVTHNQVPSLVKPEILKQTOASSISGADVAESSGSKLKIINETSQPV 65  
Db 1973 KTVALTDLSEKMEKTEQEL-----ESHQSECLHCIOVAEAEVKEKTELLQT---L 2019  
  
Qy 66 DDTVTLDFDKRTTPKPI---KDLAKGPREDKAVTENTSEKQITSGSQLEQSKES 121  
Db 2020 SDSVSELLKDKTHLQKQSLQSKLEKQALSLTKCELENQIAQLNKEKELL-----VKES 2072  
  
Qy 122 LSLINKTVPTSTNWEICD-----FITKNTLVGLS-----KSGVEKLS---QTD 161  
Db 2073 ESIQARL--SESDYEKLVNSKALEALVEKGEFALRLSSQTQEEVHQLRRGIEKLVRIEAD 2131  
  
Qy 162 -----HLV--LPSQAADGTQLQVASFATPPKKTAIABYTSRAGNGBISOLDVDG--- 211  
Db 2132 EKKQLHIAEKLERENDSL-----KDKVENLERELQWSEENQELVILDAENSKA 2182  
  
Qy 212 -----KEINEGEVENSYLKKVTIPTGYKHI-----GQ-----DAF 243  
Db 2183 EVETLTKQIEEMARSLKVFEE---LDLVTLRSEKENTKQIQKQGLSBLDKLLSFKSL 2239  
  
Qy 244 VDNKNAEVLNPESETISDYAFALHALKQIDLPDNLKAIGELAFPDNQITGKLSLPRQL 303

Db 2240 LEEKEQABEQIKEESKTAEM-----LQNLKEL-----NEAVALQCDQEI 2281  
Qy 304 MELAERAFKSNHIKTIERGRNSLKVIGEASFODNDLSQMLPDGLEKTESEFTGNPGDD 363  
Db 2282 MKATEQSLODPPTEEBEHLR-NSIEKL-RARLEADEKKQLCVLOQLKSESHHA-----D 2332  
Qy 364 HYNRRV-----VLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKN 417  
Db 2333 LLKGRVENLERELEARTNQEHAAEAENS-----KGEVETLKAKIEGWTQ--- 2378  
Qy 418 SVTGSNKGLO-----KVRKNKLEIPKQNGVTVITIGDNAPRNVDVFNKTKRYDYL 470  
Db 2379 -----SLRGLELDVVVTIRSEKENTLNEQKEQERISELEIINSFENI--LQSEKQEBKQV 2432  
Qy 471 BEVKPLSTIRKIGAFQFQSNLKSFEASDDLEIEIK-EGAFMNNRIETLELKLVTIGDA 529  
Db 2433 KEKS--STAMENLOTQLKELNERNVAALHNDQACAKQONJSSQVECELEKQAQLQGLD 2490  
Qy 530 APHINHIIAIVLPESV---QEIGRSAFRQNGANNLIPMGSKVKTILGEMAFLSNRLEHLD 585  
Db 2491 EAKNYY---IVLOSSVNGLIQEV-----EDGKQKL-----EKDEEISRLKNQIQ--- 2532  
Qy 586 LSEKQKLTPEIPQAFSDNAL--KEVLLPASLKTIRBEAFKQNHKLEVASALSHTAFNAL 644  
Db 2533 --DQQLVSKLSQVEGEHQLWKEQNL--ELRNLTVLEQKIQVLQSKNAS-----L 2579  
Qy 645 DDNDGDEQFDKNVVKVTHHNSYALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLR 704  
Db 2580 QD-----TLEVLQSSYKNLENELELTMDKMSF-----VEKVNOM-----TAK 2617  
Qy 705 QT-TQTFRDMTTAGKALLSKNSLRQGEKQKFLQEAQFPFLGRVLDLDAKAKAEKALVTKK 763  
Db 2618 ETELOREHEM--AQKTAELOEL--SGEKRLAGELQLLLEI-----K 2658  
Qy 764 ATKNGOLLERSINKAVLAYNNSAIKKA-----NVKRLKELDLTLGLVEKGPLAQATWVQ 819  
Db 2659 SSKD-QLKE-----LTLENSELKSLDOMHKDQVEKGVREETAETAYQLRHEAKKH 2710  
Qy 820 GYVLLKTPLPPEYIYGLNVYFDK-----SGKLIYALDMSDTIG 858  
Db 2711 QALLLDTN--KQYEVEIQTYREKLTSSKECLSSQKLEIDLKSSKEELNNSIKATTQIL 2767  
Qy 859 EGOKDA-----YGNPILNVDNEGYHALAVATLADYEGLDIKTILNSKLSQ 906  
Db 2768 BELKTKMDNLKYVNLKKNENARAQGMKLLIKSKCKOLE--BEKETLOKELSQLOAAQEK 2825  
Qy 907 -----TSIRQVPTAAYHRAGIFQAIQNAAAAEQQLPKPG-----THSEKSSSESAN 954  
Db 2826 QKTGTWMDTKVDELTTTEIKE---LKETLEKTKEADEYLDKYCSLLISHEKLEKAKEMLE 2882  
Qy 955 SKDRGLQSNPKTNRGRHSAIL 975  
Db 2882 TQVAHLCSQSQKQDSRGSPL 2902  
  
RESULT 63  
US-10-661-809-3  
; Sequence 3, Application US/10661809  
; Publication No. US20040101919A1  
; GENERAL INFORMATION:  
; APPLICANT: HOOK, Magnus  
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR  
; TITLE OF INVENTION: POSITIVE BACTERIA  
; FILE REFERENCE: P07741US01/BAS  
; CURRENT APPLICATION NUMBER: US/10/661,809  
; CURRENT FILING DATE: 2003-09-15  
; PRIOR APPLICATION NUMBER: 60/410303  
; PRIOR FILING DATE: 2002-09-13  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 3  
; LENGTH: 1010

```
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
us-10-661-809-3

Query Match      3.8%; Score 191; DB 16; Length 1010;
Best Local Similarity 19.9%; Pred. No. 0.0067;
Matches 215; Conservative 159; Mismatches 391; Indels 334; Gaps 50;

QY 111 SCSLEQSKESLSLNKTVPSNWEIC--DFTTK-----GNTLVGSLKSGVKEKLSQTDHL 163
DB 30 SGEKLNQSALSVKLAKQSVTAKDVQITVKGFINKGTVDGNTTV-----DQQL 77

QY 164 VLPSQAA-----DGTQLIQVASFAFPDPKKTALAEYT 195
DB 78 TIPANVAINEEKTTPSSLTLOWDQVTEATSEVERDGTGVFGNIQTNTATDFGFSLSHT 137

QY 196 SR--AGENG--EISQ-----LDVDGKEIINEGEVFNYS-----LLKQVT---IPTGY 235
DB 138 FVRVAVGKNGVSEWSEBPIKGTQDDPYKETINQVATSNLPEQGAELKLTDXDLSTGW 197

QY 236 KHIGQDAFVDNKNIAEVLNPSLETISYAFALHAK-----QID-----LPDNLKAIG 284
DB 198 -----HTWSTGIANPSDCNF--LSLKFDLGAEYQMDKIEYLPDRNAGNG 240

QY 285 EL-----APDNOITCKLSLPRQLMLAEAFKSNHIKTIEFRGN----- 324
DB 241 NILQOYRTSDGANWTFSEBPIINWKQDALTKTIETQOAYRFEVEMKVLKSVGNFGSGRE 300

QY 325 -----SLKVGEASFQNDLSQLMPLDGLKIESEAFGTG--NPGD----- 362
DB 301 MLFVKQPGTEGILHGDINDGTIDENDAMSRYNTGLSVDS--FNGYVEKGLNKGVI 359

QY 363 DHYNRVVLWTKSG----- 380
DB 360 DAYDISYVLRLQDGGIEIPDVEBEIAGGLSLAVVNBNGKDTYLPDGLTLFILKQDLKNIN 419

QY 381 GLAENTYVNPDKSIWQSPEDYTKWLEEDTYKNS-----VTGFSNKLQKVKGNK 434
DB 420 ALSTKMSDFSSFEVLGVPATNTQWENYSKYRKHSDNENLYLVLSNQG-----NK 473

QY 435 NLEIPKQHGNGVTITEIGNAFNRVDFQNKTLAKYDLEEVKLPSTIRKIGAF----- 485
DB 474 QL-----LNG-----SMDLVTPKVKVKTTRVKTATVVEQPLQFDMSQGLLVG 516

QY 486 -AFQNNLKSPASDDBEIKEGAFMNNRIETLEKDK-----LVTIGDAAPHNH----- 535
DB 517 QGFQATLSDFSVT-----VKPTLVLDKELLQALITLNOARVEKEYTPT 561

QY 536 --IYAVLPESVQEIGRSFRQN-----GANNLI FMGSKVKTIGEMAFLSNRLEHLDSEQ 589
DB 562 WAIFKPIIDEAVAVLANEQATQDVSAAENLEKASQLEKRPDVA---NKAD--LEKAIQ 617

QY 590 KOLTEIPV--QAFSDNALK--EVLPLASLKTIRBEAFKKNHLKQLEVASALSHIAFNALD 645
DB 618 EGLAKKPSDQGFETETKVLSESLAAQKVPQAEKVQEEIDQ----- 661

QY 646 DNDGDEQFNKVVVTHNSYALAGEHFIVDPDKLSSTIVDLEKILKLEGLDYSTLRQ 705
DB 662 -----ATKTLREAIQLKEQPVAVDKETLKEQIAQ--ARGRKPEGYQPT---K 705

QY 706 TTQOTOFDMTTAGKALLSKNSLRQCEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKA- 764
DB 706 ETEKQLQEAIAQAEIAIVAKETATKEVSEALNALETAMQAQKEVPLVKNQDQLEVVKRAQ 765

QY 765 ----TKNGOLLERS---INKAVLAYNNNAIKK--ANVKRLEKELDLLTGLVEGKGLAQA 815
DB 766 QVTPSEGHQFTASSIQLQELQKALLAAKNT--LKNPAANQKMWIDEAVELTSAIDG---LQEE 821

QY 816 THVQGVYLLKTPLEPIEY--IGLVNVPDKSGKLIYALDMSDTTGBEQKDAYGNPILNV 872
DB 822 VLVTDKKALEAMIAKAKIPKSPAGKEFTSESKARLTEAIDQAEGL--LADKNARQEQI--DI 879

QY 873 DEDNEGYHALAVATLADYEGLDIKTILNSKLSQLSIRQVPTAAVHRAG--IFQAIQNA 930
```

```

; Db 880 AEKN-----VKTALD--SLE-EQVLTQDTKUKKELLQKAEITLKPAGAKFTKASQDAL 929
; QY 931 AE-----AEQLLPKPGTHSEKSSSESANSK-----DRGLQSNPKTNRGRHSAILPRTG- 979
; Db 930 AEAIKQAKALVEDFNATQEAVDKCLSLQSAIEAMABEPISSNSTGNNGNHSITVSGTGV 989
; QY 980 ---SKGSFVYGIILGYTVA 995
; Db 990 TSQKGTATGTTTTKTTTS 1008

RESULT 64
US-10-282-122A-71235
; Sequence 71235, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 71235
; LENGTH: 3692
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-71235

Query Match      3.8%; Score 190.5; DB 15; Length 3692;
Best Local Similarity 19.9%; Pred. No. 0.044;
Matches 230; Conservative 184; Mismatches 399; Indels 343; Gaps 56;

QY 30 EPILKQTOASSISGADYAESGSKLKINETSGPVDVTDLPFDKXTTPEKI--KNL 87
DB 2696 ENILPATTVKDKAKADVNAKEKN--LQINS-----NDEATTEKLVASDNL 2741

QY 88 AKGPREQELKAVTS-----NTESEKQI-----TSGSOLEOS--KESLSL 124
DB 2742 -NHVETTNQAIADPDTNQVNVKNKGIGTIRDIQPLVVKKPTAKSKIESAVEKKKTEI 2800
```





QY 539 AFSDNALKEVLLPASIKTIREBAPKKNHLKQLEVASALSHIAFNALDNDGDQDFDNKV 658  
 DB 773 ---KRAVPEKL---KSLETLETFSLTLT---GNVDEE---KAL 806  
 QY 659 VKTHNSVALADGEHFIWDPDKLSSTIVDLKILKLEGLDYSLRQTQTQFRDMTTAG 718  
 DB 807 PKCE-----DRDGLSKLDKLTWVTDVRKHLDKKTD-----DPNFTG 846  
 QY 719 KALLSKSNLROGEKQKFLQEAQFFL-----GRVLDLKAIAKAEKALVTKKATKQGL 770  
 DB 847 KOIDTKTNLNRVDRFMKMKKITTLDLAKDGKIDIDITVIEAEK---KKAQDKKK 902  
 QY 771 LERSINKAVLANNASAIKANKVKLEKELDLTLGL 805  
 DB 903 EKKS-----TKKANESKVENSEIKTEL 925  
 RESULT 66  
 US-10-369-493-2070  
 ; Sequence 2070, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 2070  
 ; LENGTH: 1957  
 ; TYPE: PRT  
 ; ORGANISM: Schizosaccharomyces pombe  
 US-10-369-493-2070  
 Query Match 3.7%; Score 189.5; DB 15; Length 1957;  
 Best Local Similarity 20.6%; Pred. No. 0.021; Mismatches 451; Indels 339; Gaps 59;  
 Matches 250; Conservative 175;  
 QY 37 QASSISGADYAESSGSKLKINETSGPVDDTPT-----DUFSDKR 77  
 DB 4 QSSSGSNTSLDEESASS--LVSSAASPIDSLETPTPRNISRASTQGLAEDGDTSSQHE 61  
 QY 78 TPEKIKONLAKPREQELKAVTENTSEKQITSGSQLEBQSKESLSLKNKTPSTNSWEIC 137  
 DB 62 DSSEELKRQVRGMRHRSDLSIDAKLGSSEGSTASS-----ALPLTPRSPSNASW--- 111  
 QY 138 DPTTKGNTIVGLSKS--GVEKLSQTDHLVLPQAADGTOLIQVASFAFTPKKTAIAEY 194  
 DB 112 -LLVRG---GLDSPIDINSVTKSNLL-----NELQVRS-----KLAALAH 151  
 QY 195 TSRAGENGEIS-QLDVGDKKEIINEGEVFN-----SYLLKVTIPTGYKHIGODAFVDN 246  
 DB 152 -----ENGILSLQLSSNKKDKNTSSVTLTSEEDVSFQKLT-----NMESNFA 198  
 QY 247 KNIAEVNIPESLETISDVAFALHALKQID---LPDNLKAIGELAFPDNQITQKLSLPRLQ 303  
 DB 199 KOSEAYDLRSQLLTVE-----KLDKKEKDYKIKEDVSSI-KASLAEBQASN-KLRGEQ 252  
 QY 304 MELAERAFKSNH-IKTIEPRGNL-----KVIGE-----ASFQD 336  
 DB 253 ERLEKLLVSSNKTSTVTLQATENSRAECKTQKLEKCAINEEDSKLLEELKHNVANYS 312  
 QY 337 NDLSQLMLPDGLEKIESAFTGNPDGDDHNNRVVLWTKSGK-----NPSG-----LATEN 386  
 DB 313 AIVHKDKLIEDLSTRISF-----DNLKSERDTLSIKNEKLEKLRLNTIGSKDSRTSN 366

QY 387 TYVNPDKSLWQESPEIDYTKW-----LEEDFTYQNSVTQFSNKGLOKVKENKOLE 437  
 DB 367 SOLBEEMVELKESNRTTHSQLTDAESKLSPEQENKSLGSIDEYQNNLSSDKKKVKQVS 426  
 QY 438 IPKO-----HNGVTTIEIGDNAPRVNDFQNKTRKYDLEEVKLPSPIRKIGAPAFQSN 490  
 DB 427 SOLBEARSSLAHATGKLAEL--NSER--DFQNKIKDPEKIEQDLRACLNS-----SSN 476  
 QY 491 NLKSFEASDDLEEKKEGAPMNRITLLEKDKLVITIGDAAPHI-----NHIYAI 539  
 DB 477 ELKEKSALID---KKQDELNNLREQIKEQKVKSESTQSSLSQSLQDRLNEKKKHVEYES 532  
 QY 540 VLPESVQEIERSAFRONGANNLIWFGSKVKTIG---EMAFLSNRLHLDLSEQKOLTIP 596  
 DB 533 QLNELKGBEL-----QTEISNSELSQLSTLAAEKEAAVATNN---ELSESKNLSQTL 582  
 QY 597 VQAFSDNALKEVL-----LPASLKTIRE--EAFKKNH--LKQLEVASA----- 635  
 DB 593 CNAFOEKLAKSVMLKENEQNFSSLDTSFKKLNSHQLENNHQITITKQDKTSSKLOQL 642  
 QY 636 -LSHIAFNALDNDGDQDFDNKV-VVKTTHNSYALADGEHFIVDPDKLSSTIVDL-EKIL 692  
 DB 643 QLERANPEKESTLSDENNDLRTKLLKEESNKSILKKQE---DVSLEKNIQTLKEDLR 699  
 QY 693 KLEGLDYSL-----RQTQTQFRDMTTA-----GKALLSKSNLRQGE 731  
 DB 700 KSEEARFPEKLEAKNLREVINDLKGHETLEAQRNDLHSSLSDAKNTNAILSELTKSSE 759  
 QY 732 KQKFL-----QBA-----OFFLGRVDLDKAIK-----KASKA 758  
 DB 760 DVKELTANVETLTQDSKAMKQSFSLVNSYOSINLHLELDDHVNMQSNNTLLESESK 819  
 QY 759 LVT--KKATKNGQLLERSINKAVLAYNNSAIKKNVRELEKELD----- 801  
 DB 820 LKTDCENLTOQMTLIDNVQKLMHVNQESKSELKEVNGKLSLDLKNLSSLNVAISD 879  
 QY 802 ---LTGLVE---GKGPLAQ--ATWQGVYLLKTPPL-----PEYITGLNVYDK-SGK 846  
 DB 880 NDQILTQLAELSKNYDLSLEQSAQLNSGLKLEAKQLLHTENEELHRL-----DKLTGK 935  
 QY 847 LIYALDMSDTTIGEQKDAYGNPILNVDNDEGNYHALAVA-----TLADYEGE--DIKT 897  
 DB 936 LKIESKSSDLGK-KLTARQEEINLKEENMSQQAITSVKSLDETLSSKLEADIEH 994  
 QY 898 ILSKLSQTSIRQVPTAAHYR-----AGIFQAIONAAAEALLPKPGTHSE 945  
 DB 995 LKN-KVSEVEVERNALLASNERLMDLKNNGENTIASLQTEIEKKRAENDDLQSKLSVVSS 1053  
 QY 946 K-----SSSESANSKDRGLQSNPKT-----NRRGHSAILPRTGSKGSFVYGIIGY 991  
 DB 1054 EYENLLIISQTNKSLDKTNQLKYEKNVQKLLDEKQDRNVLEELTSK-----YKGLGE 1109  
 QY 992 TSVALLSLITAIKKK 1006  
 DB 1110 ENAQIKDELLALRKK 1124  
 RESULT 67  
 US-10-732-923-3328  
 ; Sequence 3328, Application US/10732923  
 ; Publication No. US20050108791A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edgerton, Michael D  
 ; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
 ; FILE REFERENCE: 38-15(52796)C  
 ; CURRENT APPLICATION NUMBER: US/10/732,923  
 ; CURRENT FILING DATE: 2003-12-10  
 ; PRIOR APPLICATION NUMBER: 10/310,154  
 ; PRIOR FILING DATE: 2002-12-04  
 ; NUMBER OF SEQ ID NOS: 24149  
 ; SEQ ID NO 3328  
 ; LENGTH: 1957  
 ; TYPE: PRT

; ORGANISM: Schizosaccharomyces pombe		US-10-732-923-3328	
Query Match		3.7%; Score 189.5; DB 17; Length 1957;	
Best Local Similarity		20.6%; Pred. No. 0.021;	
Matches		250; Conservative 175; Mismatches 451; Indels 339; Gaps 59;	
Qy	37	QASSISGADYAESGSKLKINETSGPVDVT	-----DLFSDKR 77
Db	4	QSSSGNTSLDDEASASS--LVSAASPIDSLETPRPNISRASTGQAEDGDTSSQHE	61
Qy	78	TYPEKIKDNLAQPRQELKAVNTESKQITSGSQESKESLKNKTVFSTSNWEIC	137
Db	62	DSEELKROEVRMRHSDLSIDAKLGSSEGSTASS-----ALPLTPRSPNASW----	111
Qy	138	DFITKNTLVGLSKS---GVEKLSQDHLVLPQAAAGTQLIQVASFAPTPDKKTAIAEY	194
Db	112	-LLVRG---GLLDSFILDINSVTQKSNLL-----NELKQVRS-----KLALEH	151
Qy	195	TSRAGENGIS-QLVDGKEIINEGEVFN-----SYLLKKVTIPTGYKHIGQDAFVDN	246
Db	152	-----ENGILSLQLSSNKKDKNTSSVTLTSEEDVSYFQKKLT-----NMESNFA	198
Qy	247	KNAEVLNLPESLETISDYAFALHAKQID---LPDNLKAIGELAFPDNQITKLSLPRQL	303
Db	199	KOSEAYDLRQLLTVE---KLDKKEKYEKIKEDVSSI-KASLAEEQASNK-SLRGQ	252
Qy	304	MLAERAFKSNH-IKTIEFRGNSL-----KVICE-----ASFOD	336
Db	253	ERLEKLLVSNKTVSTLROTENSLRAECKTLOKLEKCAINEBDSKLEELKHNVANYS	312
Qy	337	NDLSQLMLPDGLEKIESEAFTPGNPDHNNRVLMTKSGK-----NPSG-----LATEN	386
Db	313	AIVHKDLIEDLSTRISF-----DNLKSERDTLSIKNEKLEKLRTIGSKDSRTSN	366
Qy	387	TYNPDKSLWQSPEDITYKW-----LEEDFTYQKNSVTGFSNKGLOKVKRNKULE	437
Db	367	SQLEEEVMELKESNRTHISQLTDAESKLSFQENKSLGSDIYQNNLSKDKMKVQVS	426
Qy	438	IPKQ-----HNGVTITEGDAFRNVDPFQNTLAKYDLEVKLPSTIRKIGAFQAFON	490
Db	427	SQLEEARSLAHATGKLAEI--NSER--DFQNKIKDFEKIEQDLRACLNS-----SSN	476
Qy	491	NLKSFEASDLEBIEKEGAFMNNRIETLEKDKLVITIGDAAFHI-----NHIYAI	539
Db	477	ELKEKSALID---KKQQLNNLRQIQKEQKVSESTQSSLOSQDLQILNEKKKHEVYES	532
Qy	540	VLPEVQETGRSAFRONGANNLIFMGSKVKTIG---EMAFLSNRLHLDLSQKQLTEIP	596
Db	533	QLNELKGL-----QTEISNGEHLSSQLSTLAAEKEAAVATNN-----ELSEKNSLQTL	582
Qy	597	VOAFSDNALKEVL-----LPASLKTIRE--EAFKQNH--LKOLEVASA-----	635
Db	583	CNAFQEKLAQVNLKENSQNFSSLDTSFKKLNSHQLNENHNTTQKLDKDTSSKLQOL	642
Qy	636	-LSHTAFNALDNDGDGEQDNKV-VVKTTHNSVALADGEHFTVDPKLSSTIVDL-EKIL	692
Db	643	QLERANFEQKESTLSDENNDLRTKLLKLBESKSLIKQOB---DVDLSEKNQTLKEDLR	699
Qy	693	KLIEGLDYSTL-----RQTQTFQDMTTA-----GKALLSKSNLRQOE	731
Db	700	KSHEALRFKLEAKNLREVIDNLKGKHETLEAQRLDHSLSLDAKNTNAILSELTKSSE	759
Qy	732	KQKFL-----QEA-----QFLGRVDLDKAIA-----KAEKA	758
Db	760	DVKRLTANVETITQDSKAMKQSFSLVNSYQISNLHYELRHHVNMQSNNTLSESK	819
Qy	759	LVT--KKATKNGQLLERSINKAVLAVNSAISKANVKRLKELDL-----	801
Db	820	LKTDENLTQONTLLIDNVQKLMKHVNOESKVELKEVNGKLSLDLKNLRSNLVAISD	879
Qy	802	----LTLGLVE---GKGPLAQ---ATMVQGVYLLKTPLP-----PEYIGLVYFDK-SGK	846

Db	880	NDQILTLAELSKNYDSLEQESAQLNSGLKSLAEKQLLHTNEELHRL-----DKLTGK	935
Qy	847	LIVALDMSDTTIGEGOKDAYGNPILNVDBDNEGYHALAVA-----TLADYBGL--DIKT	897
Db	936	LKIESKSDLGK-KLTARQBEISNLKEENMSQQAITSVKSKLDETISKSLLEADIEH	994
Qy	898	ILNSKLSQLTSTRQVPTAAYHR-----AGIFQAIONAAAEAFQLLPKPTHSE	945
Db	995	LKN-KVSEVEVERNALLASNERLMDLKNNGENIASLQTEIEKKRAENDDLQSKLSVSS	1053
Qy	946	K-----SSSESANSKDRGLQSNPKT-----NRGHSAILPRTGSKGSFVVCILGY	991
Db	1054	EYENLLLSQNTKSLDKTNQIKYIEKNVQKLDKQDNVLEBELTSK-----YKLCGE	1109
Qy	992	TSVALLSLITAIKKK 1006	
Db	1110	ENAIQKDELLALRKK 1124	
RESULT 68			
US-10-282-122A-70176			
; Sequence 70176, Application US/10282122A			
; Publication No. US20040029129A1			
; GENERAL INFORMATION:			
; APPLICANT: Wang, Liangsu			
; APPLICANT: Zamudio, Carlos			
; APPLICANT: Malone, Cheryl			
; APPLICANT: Haselbeck, Robert			
; APPLICANT: Ohlsen, Kari			
; APPLICANT: Zyskind, Judith			
; APPLICANT: Wall, Daniel			
; APPLICANT: Trawick, John			
; APPLICANT: Carr, Grant			
; APPLICANT: Yamamoto, Robert			
; APPLICANT: Forsyth, R.			
; APPLICANT: Xu, H.			
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms			
; FILE REFERENCE: ELITRA.034A			
; CURRENT APPLICATION NUMBER: US/10/282,122A			
; CURRENT FILING DATE: 2003-02-20			
; PRIOR APPLICATION NUMBER: 60/191,078			
; PRIOR FILING DATE: 2000-03-21			
; PRIOR APPLICATION NUMBER: 60/206,848			
; PRIOR FILING DATE: 2000-05-23			
; PRIOR APPLICATION NUMBER: 60/207,727			
; PRIOR FILING DATE: 2000-05-26			
; PRIOR APPLICATION NUMBER: 60/230,335			
; PRIOR FILING DATE: 2000-09-06			
; PRIOR APPLICATION NUMBER: 60/230,347			
; PRIOR FILING DATE: 2000-09-09			
; PRIOR APPLICATION NUMBER: 60/242,578			
; PRIOR FILING DATE: 2000-10-23			
; PRIOR APPLICATION NUMBER: 60/253,625			
; PRIOR FILING DATE: 2000-11-27			
; PRIOR APPLICATION NUMBER: 60/257,931			
; PRIOR FILING DATE: 2000-12-22			
; PRIOR APPLICATION NUMBER: 60/267,636			
; PRIOR FILING DATE: 2001-02-09			
; PRIOR APPLICATION NUMBER: 60/269,308			
; PRIOR FILING DATE: 2001-02-16			
; Remaining Prior Application data removed - See File Wrapper or PALM.			
; NUMBER OF SEQ ID NOS: 78614			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 70176			
; LENGTH: 2398			
; TYPE: PR			
; ORGANISM: Staphylococcus aureus			
; FEATURE:			
; NAME/KEY: MISC FEATURE			
; LOCATION: (311)..(311)			
; OTHER INFORMATION: X=any amino acid			
; FEATURE:			
; NAME/KEY: MISC FEATURE			

```
; LOCATION: (313)..(313)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (327)..(327)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (328)..(328)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (336)..(336)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (350)..(350)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (362)..(362)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-70176

Query Match      3.7%; Score 189; DB 15; Length 2398;
Best Local Similarity 18.7%; Pred. No. 0.03;
Matches 210; Conservative 153; Mismatches 444; Indels 314; Gaps 46;

QY 2 KKHKTVALTITVSVVTHN-----QEVFSIVKEPIIKQTOASSISGADYAESGSKS 55
DB 1111 KQHLG---TLTHITTAQRNDLTNQISQATNLGAVESYKQN--ANSLDGA-----MGNLQ 1159

QY 56 LKINETSGBV-----DDTVTDLPFSKRTTPPEKID-----NLAKGREBELKAV--- 99
DB 1160 TAINDKSGTLASQNFILDADQKRNAYNOAVSAETILNKQTPNTAKTAVEQALANNVNA 1219

QY 100 -----TENTESEKQ-----ITSGSOLEQ-SKESL--SINKTVPSTSNWEICDFTKGNT 145
DB 1220 KHALNGTONLNNAKQAATAINGASDLNQKQDALKQAQANGAQRVSNADQVOHNAETLNT 1279

QY 146 LVGLSKSGV-----EKLSDTDHLV-----LPSQAADGT 173
DB 1280 ANGALKHAIADKNTNLASSKYVNAUSTKQNAVTTKVTNAEHIISGTPVTVPSTVETAA 1339

QY 174 QLIQVASFAPTD-----KKTAAEYTSRAGENGESIQOLDVD 210
DB 1340 NOVNSAKQELNGDERLREAKQANATAIDALTQLTNPQAKLKEQVQANRLEDVQVTN 1399

QY 211 GKEINEGEVFNYSLLKKVTIPTGYKHIGQDAPVDNKN-----IAEVNLPESL 258
DB 1400 GOALNNAMKGLRDSIANETVKTSONY--TDASPNQSTYNSAVSNAKGIINQTNPN--- 1454

QY 259 ETIDYAPFAHLAKQIDLPDLNKAIGELAFFONQITGKLSL-----PRLMLRAERAF 311
DB 1455 -TMDTSAITQATOVNNAKNGLGAENLR--NAQNTAKONLTLHLTNQKSAISSQID 1511

QY 312 KSNHKTIFRONSUKVIGEASFQNDLSQMLPDGLEKIESEAPT--GNPGDDHYNNRV 369
DB 1512 RAGHVSEVATKNAAT---ELNTQGNLEQAIHQDQNTVK-QSVKFTDADKAKRDATYNAV 1567

QY 370 -----VLWTSGKNPSGLATENTYNNPDKSLWQESPEIDYTKWLBEDFTYQKNSVTGFSN 424
DB 1568 SRAEAILNKTQAGNTSKDVEAAIQNVSSA-----KNALNGDON 1606

QY 425 KGLQKVRKNKLEIPKQHGVTITTEIGDNAPRNVDFQNKTLRKYLEEVKLPSTIRKIGA 484
DB 1607 -----VTNAKN-----AAKVALNLTISINNAQKEDLTTKIDQATTAVGEA 1647

QY 485 FAPQSNLKFBRASDLEIKGAFMNNRIETLEK-----DKLVTTIGDAFHNHIY 537
DB 1648 VNSTQLNTAMAN-----LONG--INDKNTLASENYHDADSCKKTATYQAVTNANIL 1700

QY 538 -----AIVLPESVQIGRSAPRQNGANNLIPMGSKVKTLGEMAFLSNRLHLDLSE 588
```

```
DB 1701 NMSGSLDKTAVENALSQVANAKGALNGHNLEQAKSNANT-----TINGLOHLTTAQ 1754
QY 589 QKQLTEIPVQAFSDNALKEVLLPASL-----KTIREAFKXNHLKQLEVASALS 639
DB 1755 KDKLKQVQQAQNVAGVDTVKSSANTLNGAMGTLRNSTQDNTATKNGQVLDATERNKTN 1814
QY 640 AFNALDDNDG-----DEQFDNKVVVKTHNSVALADGHEFIVDPDKLSSTIVD-- 687
DB 1815 YNNAVDSANGVINATSNPNMDANAINQIATQVTSKQNAL-DGTNLTQAKQATNAIDGA 1873
QY 688 --LEKILK-----LIEGLD-----YSTL 703
DB 1874 TNLNKAQDALKAQVTSQQRVANVTSIQQTANELTANGLOHQGIDDENATKQTKYRDA 1933
QY 704 RQTTOTQFRDMTTAGKALLSK---SNLRQSKQKPLQEA-----OFFLGRVLDKAIKAE 756
DB 1934 EQSKTAYDQAVAAAKAILNKQTSNSDKAAVDRALQOVVTSKDALNGDAKLAEKAAK 1993
QY 757 KALVTKKATKNGQL--LERSINKAVLAYNNSAIKANVVRLEKELDLTLGLVEGKGLAQ 814
DB 1994 QNLGLNLHITNAQRDLEGGQINQATTVDGVNTV-KTNANTLDGAWNSLQGSINDK---- 2048
QY 815 ATMVQGVVLLKTPPLPEYIYGLNVYFDKSGKLIYALDMSDTIGEGQKDAYCNPILNVDE 874
DB 2049 ATLNRQNYL-----DADESKBNAYTQAVTAAEGILNKQTCGN-TSKADV 2091
QY 875 DNEGVALAVATLAD--YEGLDIKTILNSKLSQVTSIRQVP 913
DB 2092 DN-----ALNAVTRAKAALNGAD--NLNNAKTSATNTIDGLP 2126
```

## RESULT 69

```
US-10-732-923-3332
; Sequence 3332, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3332
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-732-923-3332
```

```
Query Match      3.7%; Score 188.5; DB 17; Length 1268;
Best Local Similarity 21.2%; Pred. No. 0.013;
Matches 229; Conservative 160; Mismatches 393; Indels 297; Gaps 54;

QY 94 QELKAVTENSEKQITSGSQLEBSKESLSLNKIVP---STS-----N 133
DB 9 BELREITRN-----VTTGNDLE-DDEPLKAIQITSELLTSLTAADIRIPISYLTFLIY 61
QY 134 WEICDFITKGNITLVLSKSGVEKLSQTDHLVLPQAAD-----GTQLIOVA---SPAPT 184
DB 62 WLFQDFKATNDFL--SDKSVIKSLLSFSYQI--QDEVTIKCLVTLMLGVAYEFSKES 116
QY 185 PDKKTAIAEY-TSRAGEN---GEISQLDVD--GKEIINEGEV-----PNS 223
DB 117 PPRKEYPEFITKTLKDNVYASRIKQPKKDSVFSKVDNEDSILTPDELDTGLPKVYFST 176
QY 224 YLLKVT-----IPGYKHIGQDAPVDNKNIAEYNLPE-----SLETTIDYAF 267
DB 177 YPIQLFENIYRITRALTSHDPEBPINKISPEEVEKQROCTKLGKETSQTETESTHE 236
QY 268 HLAKQIDLPNLKAIAGELAFPDN-----QITGKLSLPRQLMLRAER 309
```

```
Db 237 NLTEKIALTNEHEKDEKQIILNSHSHSUKENFISILETELKXVRSLDMDTQLRDVLET 296
Qy 310 AFSKNIHKTIEPRGNSLKVIGEASFQNDLSQLMLPDGLEKIESEAFNGPGDDHNNRV 369
Db 297 KOKENQALLEYKSTHK-----QEDSIK--TLEKGLTILSQK---KXEDGIN--- 341
Qy 370 VMTKSGKPNPGLATENTYVNP--KSLWQESPIDY-----TKWLEEDFTYQKNSVTGFS 423
Db 342 ----KMGKOLFALSRQWAEVENCNKLQEKDKSNVNHQKETSLSKEDIAAKITEIKAI- 396
Qy 424 NKGLQKVK-RNKULEIPKOHNGVTIITEIGD-----NAFRNVDFQW 462
Db 397 NENLEEMKIQCNNLSEKHEKISHELVEYKSRFQSHDNVLAKTEKLKSLANNYKQMAEN 456
Qy 463 KTLRKYDLLEVKLPSTIR-----KIGAFAPQSNLKL-----SPEAS-----DLEBI 504
Db 457 ESILIK-AVESKNESSIQNLQNKIDMSQEKENFQIERGSEKIEKIEBQKKTISLEQT 515
Qy 505 KEGAFMNNRIETLE-----LKDKLVTIGDA-AFHINHIYAIIVLPESVQIEGRSAFRON 556
Db 516 KEBIISKSDSSKDEYESQISLLKEKETATTANDENVNKISELTKTRELEAEALAYK-N 574
Qy 557 GANNL-IFMGSKVKTIGEMAFNLNRLEHLDLSEKQKLTPELQVAFSDNALKEVL----- 609
Db 575 LKNELETKLETSEKALKVEKNEHEHLKEEIKIQLEKEATETKQQLNSLRANLESLEKEHEB 634
Qy 610 LPASLKTIREEAPKNHLKQLEVASALSHTAFNALDDNDGDEOFDNKVVVKTHNSYALA 669
Db 635 LAAQLKKBEOIANKERQYNEEISQ-----LNDEITSTQOENESIKKKNDE-----L 681
Qy 670 DGHFTIV-----DPDKLSSTIVDLEK-----ILKLEIGLDYSTL----- 703
Db 682 EGBVKAAMKSTSEBQSNLKKSEIDALNLQIKELKKNETNEASLESIKSVESETVKIKEL 741
Qy 704 -----RQTQTQFRMDTAGKALLSKNSLRQEKOKFLQ-----EAOFFLGRV-DLD 749
Db 742 QDECNFKEKEVSELDKASEDKNSKYLELOKESEKIEELDAKTELKIQLEKITNLS 801
Qy 750 KAKAEKALVTYKAT-----KNG-QLLERSINKAVLAYNNSAIKANKVRLKELDLIT 803
Db 802 KAKESESELRLKKTTSBERKNAEBOLEKLNKNE--IQIKNQ-----FEKERKLIN 851
Qy 804 GLVEGKPLAQATWQGVYLLTKPLPEYIIGLVYFDKSGKL-----IYALDMSDTIG 858
Db 852 ---EGSSTIQ-----EYSEKINTLEBELRLQNELKAKEIDNTRS 891
Qy 859 EGQKDAYGNPILNVDENEGYPHALAVATLADYGLDIKTILNSKLSQTSI-----RQVP 913
Db 892 ELEKVSLSNDELLEEKQ-----TIKSLQD-EILSYKDKITRDNDEKLLSIERNKRDLE 944
Qy 914 TAAYHRAGIFOATONAAAEQOLLPKPGTHSEKSSSSSESA---NSKD--RGLQSNPKTN 967
Db 945 SLKEQ-----LRAAQESKAKVEELKK-----LEBESSKEAELEKSKMMKKLESTIESN 995

RESULT 70
US-10-282-122A-60608
; Sequence 60608, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
```

```
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60608
; LENGTH: 2013
; TYPE: PRF
; ORGANISM: Listeria monocytogenes
; US-10-282-122A-60608

Query Match 3.7%; Score 188.5; DB 15; Length 2013;
Best Local Similarity 21.9%; Pred. No. 0.025;
Matches 250; Conservative 148; Mismatches 429; Indels 315; Gaps 66;

Qy 18 VTHNEQVSLVKEPILKOTQASSISGADYABSSGSKLKNETSGPVDDT--VTDL--- 72
Db 975 ITLFEAIVPTDAPRDEIAYNSIAYRIDKETGSGTSLASEPPRGGVKSTPSTDLNTA 1034
Qy 73 ---FSDKRTTPBKIKDNLAKPREQLKAV-----TENTESEK---QITSGSQLESKE 120
Db 1035 GNSFSD-----LNKNGVKDADELGLNAVKLDLYKONNEPEKVEVTYSSDALDKKG 1087
Qy 121 SLSLN-----KTVPSTSNWEICDPIYKG-NTLVGLSKSG-----VEKLSQTDHLVLP 168
Db 1088 LPDFNGLNNGTYKIAAHLFN-KNADFITTPGNKIVKDSKDDSIGHITVNNSTFTI--DD 1144
Qy 169 AADGTQL---IQVASFAFTP-----DKKTAIAEYTSRAGENGESIQLDVDGKEI 214
Db 1145 LANGNPKNLVGLQIPIYATPIKGAVFVNKDGEPKITSY-----GQDYHIALFDKDGKEV 1200
Qy 215 -----INEGEV-FNSYLLK-----KVTIPTGYKHIGQDAFVNDKNIAEYNLPESLET 260
Db 1201 QSAIKTNSKGPSFNDVAIKNPADPKLVTAFTGTFNV-----YSAKN----- 1243
Qy 261 ISDYAFALHALKQIDLPDNLKAIAGELAFDQNOITKGLSLPQMLRAEAFKSNHKTIE 320
Db 1244 ----PLFNMSTKEYTLNSVPGVGGA--EYIYT-ETSKPTTKI-ILDKAVTPNAI-TIE 1294
Qy 321 PRGNSLKIVGEASFQNDLSQLMLPDGLEKIESEAFNG-----PGDDHNNRVVLTWTKSG 376
Db 1295 SSDEATEVTNEWTLSDSGTGVVY-----TGTGNTIRIPNDE-----GTIYA 1335
Qy 377 KNPSGLATENTYVNPDKSLWQESPIDY---TKWLEEDFTYQKNSVTGFSNKGLOKVRKN 433
Db 1336 KNTATDEAGNT-ASDEKTF-----DIDYTVPLTVNQDASAEVNSTE--ANIGWK--P 1384
Qy 434 KNLKIPKOHNG-----VTITEIGDNA-----FRNVDFQWKT- 464
Db 1385 LNVAATDTHDGNITPVDVYSKVKDVLGTPVTVTATDASGNKATQTNLRIVDTTSPIT 1444
Qy 465 -----LRKYDLEEVKLPSTIRKICA-----FAFQSNLKL 494
```

```
Db 1445 LITNPLTYSIENMRK--LTEQELTYAAGLIGDNDYDLAPGAQVQPNKQPMVTFSTI 1502
Qy 495 FRASDDLEIEKGAFM-----NNRIE---TLEKDKLVLTIGDAAFHINHIIYAIVL 541
Db 1503 F---GDIAVKAQGVQVQVNLADSSGQAIPTITINVVD---TVGPVKAADVSHVNT 1556
Qy 542 PESVOEIGBSAFRONGANNLIIFWGSVKVTLGEMAFSLNRLHLDLSE--QKOLTEIPVQA 599
Db 1557 TKTEAEFFDARLDVTDNN-----DDTDLIITSNFAEKVNLNPKGVEVITATDT 1608
Qy 600 PSDNALKEVLLPASL-KTIREBAFKKNHLKQLEVASA--LSHIAFNAALDDNDGBOFNK 656
Db 1609 KGNQTTKEITVQVSKDKPVITADPKISYQKIEVTEANFSLGVAHEVTDLDGDKVITSD 1668
Qy 657 VVVKTHN---SYALA---DGEHFIVDPDKLSSITVDLEKILKILIEGLDYSTLRQTQT 709
Db 1669 FAEKVDFNKVGTYYTTLNAKDEYGNATBEPKVSYSI--FNKIAPTFFNADNKTIEAVNEL 1726
Qy 710 QPRD--MTTAGKALLSKSNLRQGEKQFLOBAQFFLGRVLDLKAIAKAEKALVTKKATK- 766
Db 1727 PLSIESIFKIEADYLSGNKL-----KVYTPPE-QTIKGNVPGEYSI-----KVYTKD 1772
Qy 767 -NGQLLERSINKAVLAYNNSAIKKA-NVKRLB---KELDLL-----TGLVEG---KGP 811
Db 1773 DSGNTAEVTVITIKDTGPSIMTKSTYKLEVSQKEPNWITFFGIKATDIVDGDVTKNI 1832
Qy 812 LAQAATWQV---GVYLLKTPLEPYIYIGLVNLYFDKSGKLIYALDMSDTIGSGQDAYGN 867
Db 1833 KVDSEVNLNKVGTY-----PIYFTVTDALGNESKLSKSTVQIVDT-----SS 1875
Qy 868 PILAVDEDEGHALAVATLADYEGI-DIKT-ILNS--KLSQTSIRQVPTAAHYHRAGIF 923
Db 1876 PELTIDKLEISYP--TGKTSIDQFLQDITGKVTNSYGTGVKVTNNLSKI--VDMNKAGKY 1931
Qy 924 QAIQNA-----AABAEOLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPR 977
Db 1932 KVTVTATNSSGGVAEKTILL-----TVKNSDSSFIAPVSKD-----DNKKPKAKNIPK 1979
Qy 978 TG 979
Db 1980 TG 1981

RESULT 71
US-09-815-242-5816
; Sequence 5816, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
```

```
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5816
; LENGTH: 2478
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5816

Query Match 3.7%; Score 187.5; DB 9; Length 2478;
Best Local Similarity 19.3%; Pred. No. 0.039;
Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;

Qy 12 LTTVVVTHNOEVFSLVKEPILKQTOASSSISGA-----DYAESSGSKLKINETSGPV 65
Db 1170 ITAEIEAANADVNAV-----TQANSNIEAANSQNDVDQAKTTGENSI----- 1213
Qy 66 DDTVTDLPFSDIKRTPPEKIKONLAGPBEQELKAVTENTESEKQITSGSQLESLSLN 125
Db 1214 -DQVTPTVNKKATARNETAIL--NNKLEIQATPDATDEEKQ---AADAENTENGKAN 1267
Qy 126 KTVPSTSNWEICDFTTKGNTLVGLSKSGVKLSQTDHLVLPSQAADGTQLLIQVASFATP 185
Db 1268 QAISAAAT-----NAQVDEAKANAE-----AALNAVTP 1295
Qy 186 D--KRTAIAEYTSRAGENGESQLDVGKEIINEGEVFNLSYLLKKVVTIPTYGKHIGQDAF 243
Db 1296 KVVKKQAAD-----EIDQLQATQTVINN-----DQNAI 1325
Qy 244 VDNKNIAEVLNPELSLETISDVAFPHALAKQIDLDPNLKAIGELAFPNQITGKLSLPRQL 303
Db 1326 TEKK-----EAAIQQLATAVTDKANNITA---ATDDNGV----- 1356
Qy 304 MRLAERAFKSNHIKTIBFRGNLSKVIGEASQDNDLSQLMLPDGLEKIESEAFNGNPGDD 363
Db 1357 -----DQAKDAGKNSIQSTQPATAVKSNKANDVDQAVTTQNOAID 1396
Qy 364 HYNRRVVLWTKSGKGNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFS 423
Db 1397 -----NTTGATTEK--NAAKDLVLKAKEKAYQDILNAQT---NDVTQIK 1437
Qy 424 NKGLQKVRKNKLEIPKQHNQV-----TITEIGDNAPRNVDPONKTL--RKVDLEEVKLPS 477
Db 1438 DQAVADIQ-----GITADTTIKVDKDELATKANEQKALIAQTADATTEKEQ 1485
Qy 478 TIRKIGAPAFQSN-NLKSFEASDDLBEIKEGAFM-----NNRIETL-ELKDK 522
Db 1486 ANQVVDALQTOGNQNIENNAQSIDDVNTAKONAIQAIIDPIQASTDVTKNARAEIJTEMQNK 1545
Qy 523 LVITIGDAAFHINHIIYAIVLPESVOEIG--RSAPFRQNGANNLIIFWGSVKVTLGEMAFSLNR 580
Db 1546 ITEILNNNETTNE-----EKGNDIGPVRAAY--EGLNNI-----NAATTG----- 1585
Qy 581 LEHLDLSEKQLTEIPVQAFPSDNALKEVLLPASLKTIREBAF-KKNHLKQL----- 630
Db 1586 ----DVTTAKDQVAVQVQQLHANPVKK---PAGKKELDQAAADKKTQTEQTPNASQOEIN 1638
Qy 631 ----EVASALSHIAFNALDDNDGDEQFDNKV-----VVKT--HNSYALADGEHF-- 674
Db 1639 DAKQEVDTLNOAKTN--VDQSTWNEYVDNAVKEGAKINAVKTFSEYKDKALAKIEDAYN 1697
Qy 675 --IVDDP-----KLSSTIVDL-EKILKILIEGLDYSTLRQTQ-----TQFRDMT- 715
Db 1698 AKVNEADNSNASTSSEIAEAKQKLAELKQTDQNVNQATSKDDIEVOIHNDLDINDYTI 1757
Qy 716 TAGKA-----LLSKSNLRQGEKQFEL-----QEAQFPLGRV-----D 747
Db 1758 PTGKESATTDLYAYADQKKNINISADTNA TODEKQAIKQVDQNVQTALESINNGVDNGD 1817
Qy 748 LDKAIAKAEKALVTKKATNGQLLERSINKAVLAYNNSAIKKNVVKRLEKELDLITG--- 804
```

Db 1818 VDDALTQKAAI---DAIQVDAATVKPKANQAI-----EVKAEDTKESIDQSQDLTAEEK 1868  
QY 805 -----LVREKGPPLAQATMVQGVYLLKTLPLPEYVIGL---NVVFKSGK--- 846  
Db 1869 TEALAMIKQITDOAKQGITDATTAEVEKAKAQ-----GLEAFDNQIDSTEKQKAI 1920  
QY 847 --LIYALDM-----SDTIGGQKDAYGNPILNVDENEGYPHALAVATLADY-EGLDI 895  
Db 1921 BELETALDQIEAGVNVNADATTE-EKEAFTNAL-----EDILSKATEDISDQTTNAEI 1972  
QY 896 KTLNSKLSOLTSIROVPTAAVHRAGI-----FOAIONAAAE----- 934  
Db 1973 ATVKNBALEQLKQRIINPEVKKNALREIVVKNQIEIKKNADADASAKIARTDLGRYF 2032  
QY 935 -----QLLPKPGTHSEKS-----SSSESANSKDRGLOSQPKTNRGRHSAIL 975  
Db 2033 DRFADKLDKTTQTNVAEVLQNVTPAIEAIVPQNDPDANDTNGDNDNATNSANATP 2092  
QY 976 PRGTG 979  
Db 2093 ENTG 2096

## RESULT 72

US-09-815-242-12967  
; Sequence 12967, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Onisen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12967  
; LENGTH: 2478  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12967

Query Match 3.7%; Score 187.5; DB 9; Length 2478;  
Best Local Similarity 19.3%; Pred. No. 0.039;  
Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;  
QY 12 LTTVSVVTHNOEVSILVKEPILKQTCASSISGA-----DVAESSGSKLKINETSGPV 65  
Db 1170 ITAEETAANADVNAV-----TQNSNIEANSQNDVDQAKTTGENSI----- 1213  
QY 66 DDTVTDLFSDKRTTPKPKIDNLAQPREQELKAVTENTSEKQITSGSQLEQSKESLSLN 125

Db 1214 -DQVTPTVNNKATARNEITAIL--NNKLOEIQATPDATDEEQ---AADAEBANTENGKAN 1267  
QY 126 KVPSTSNWEICDFITKGNLTGLSKSGVGLKSLQTDHLVLPQAADGTQLIQVASFATP 185  
Db 1268 QAIASATT-----NAQVDEKANA-----AANAVATP 1295  
QY 186 D--KKTATAYTSRAGEGEISQLDVDGKEIINEGEVFNYSYLLKKVTIPTGYKHQQDAF 243  
Db 1296 KVVKQQAQKND-----EIDQLQATQTNVNN-----DQAT 1325  
QY 244 VDNKQIAEVLNPLESLETISDYAFALHALQIDLPDLNKLKAGELAFDNOITGKLSLPROL 303  
Db 1326 TEEK-----EAAIQQLATAVTAKNNITA-----ATDDNGV----- 1356  
QY 304 MELAERAPKSNHIKTIEPRGNSLKVIGEASFQNDLSQMLPDGLEKIESEAFNGPGDD 363  
Db 1357 -----DOAKAGKNSIOSTQPATAVKSNKNDVDQAVTTQQAID 1396  
QY 364 HYNRRVWLWTKSGKNPSGLATENTYVNDPKSLWQESPEIDYTKWLEEDFTYKNSVTGFS 423  
Db 1397 -----NTTGATTEK--NAAKDLVLKAKEKAYQDILNAQTT---NDVTQIK 1437  
QY 424 NKGLOKVKKNKLEIPKQHNGV-----TTEIGDNAPRVNDPQNTL--RKYDLEEVKLS 477  
Db 1438 DOAVADIQ-----GITADTTIKDVAKDELATKANEQKALIAQTADATTEKEQ 1485  
QY 478 TIRKIGAFAFQSN-NLKSFEASDDLEEIKEGAFM-----NNRIETL-ELKDK 522  
Db 1486 ANQVQDAQLTOGNQNIENQAQSIDDVNTAKDNAIQIDPIQASTDVKTNARAEILTEMQNK 1545  
QY 523 LVTIGDAAFHINHIYAIVLPEQVQIG--RSAFRONGANNLIFMGSKVKVTLGEMAFLSNR 580  
Db 1546 ITEILNNNETTNE-----EKGNDIGPVRAAY-EEGLNNI---NAAATTTG----- 1585  
QY 581 LEHLDSQKQLTEIPVQAFPSDNALKEVLLPASLTIREAF-KGNHLKQL----- 630  
Db 1586 ----DVTTAKDTAVQVQQLHANPVKK---PAGKELDQAAADKKTQIEQTPNASQOEIN 1638  
QY 631 ----EVSALASHIAFNALDDNDGDQFQNKV-----VVKT--HNSYALADGEHF-- 674  
Db 1639 DAKQEVDTLNOAKTN--VDQSTNEYVDNAVEKGKAKINAVTPEYKDKALAKIEDAYN 1697  
QY 675 --IVDPD---KLSSTIVDL-EKILKLEGLDYSLRTQTT-----TQFRDMT- 715  
Db 1698 AKVNEADNSNASTSSEIAEAKQKLAELKQTDQNVNQATSKDDIEVQIHNDLNDINDYTI 1757  
QY 716 TAGKA-----LLSKSNLRQSEKQKFL---QEAQFFLGRV-----D 747  
Db 1758 PTGKESATTDLYAYADQKKNISADTNATQDEKQQAIKQVDPQNVQTALESINNGVDNGD 1817  
QY 748 LDKATAKAEKALVTKKATKNGQLLERSINKAVLAYNNSAIKANKVKLEKELDILTG--- 804  
Db 1818 VDDALTQKAAI---DAIQVDAATVKPKANQAI-----EVKAEDTKESIDQSQDLTAEEK 1868  
QY 805 -----LVEGKGLAQATMVQGVYLLKTLPLPEYVIGL---NVVFKSGK--- 846  
Db 1869 TEALAMIKQITDOAKQGITDATTAEVEKAKAQ-----GLEAFDNQIDSTEKQKAI 1920  
QY 847 --LIYALDM-----SDTIGGQKDAYGNPILNVDENEGYPHALAVATLADY-EGLDI 895  
Db 1921 BELETALDQIEAGVNVNADATTE-EKEAFTNAL-----EDILSKATEDISDQTTNAEI 1972  
QY 896 KTLNSKLSOLTSIROVPTAAVHRAGI-----FOAIONAAAE----- 934  
Db 1973 ATVKNBALEQLKQRIINPEVKKNALREIVVKNQIEIKKNADADASAKIARTDLGRYF 2032  
QY 935 -----QLLPKPGTHSEKS-----SSSESANSKDRGLOSQPKTNRGRHSAIL 975  
Db 2033 DRFADKLDKTTQTNVAEVLQNVTPAIEAIVPQNDPDANDTNGDNDNATNSANATP 2092  
QY 976 PRGTG 979  
Db 2093 ENTG 2096



```
RESULT 73
US-10-470-048B-220
; Sequence 220, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; FILE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN
; FILE REFERENCE: SONN:035US
; CURRENT APPLICATION NUMBER: US/10/470,048B
; CURRENT FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 220
; LENGTH: 2478
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-10-470-048B-220

Query Match          3.7%; Score 187.5; DB 17; Length 2478;
Best Local Similarity 19.3%; Pred. No. 0.039;
Matches 221; Conservative 150; Mismatches 380; Indels 393; Gaps 54;

QY      12 LTTVSVTHNQSVFLVKEPILKQTOASSISGA-----DYAESSGSKKLKINETSFPV 65
DB      1170 ITABEIAAANADVDAV-----TQANSIEAANSQNDVDQAKTTGENSI----- 1213

QY      66 DDTVTDLFSDKRTTEPEKIDNLAKPREQELKAVTENTSEKQITSGSQLEOKESLSLN 125
DB      1214 -DQVPTVNNKATARNETAIL--NNKLEIQATPDATDEKQ---AADAENANTENGKAN 1267

QY      126 KTVPTSNNWEICDFITKGNLTVLGSKGVKLSQTDHLVLPQQAADGTQLIQVAFATP 185
DB      1268 QAISAATT-----NAQVDEAKANAE-----AANAVTP 1295

QY      186 D--KKTATAYTSRAGENGIEISQVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAF 243
DB      1296 KVVKQAQAKD-----EIDQLQATQTNVNN-----DONAT 1325

QY      244 VQNKIAEVLNPLESLETISDYAFALHAKQIDLPNLKAIGELAPFDNQITGKLSLPRQL 303
DB      1326 TEK-----EAAIQLATAVTAKNNITA-----ATDDNGV----- 1356

QY      304 MRLAERAFKSNHIKTIEPRGNSLVIGESFQNDLSQLMLPDGLEKIESEAFPTGNPGDD 363
DB      1357 -----DQAKDAGKNSIQSTQPATAVKSNKNDVDQAVTTQQAID 1396

QY      364 HYNRRVVLTKSGKPSGLATENTVYNPKSLWQESPEIDYTKWLEEDFTYQKNSVTGFS 423
DB      1397 -----NTTGATTEK--NAAKOLVLKAKKAYQDILNAQTT---NDVTQIK 1437

QY      424 NKGLOKVKRNKNLEIPKQHNGV---TITEIGNAPRNVDFQNKTL--RKYDLEEVKLPS 477
DB      1438 DOAVADIO-----GITADTTIKDVAKDELATKANEQKALIAQTADATTEEKQ 1485

QY      478 TIRKIGAFAPQSN-NLKSFEASDDLEIEKEGAPM-----NNRIETL-ELKDK 522
DB      1486 ANQVDAQLTQGNQNIENASIDVVNTAKDNAIQDIPQASTDVKTNARAEELLTEMQNK 1545

QY      523 LVITGDAAPHIHIIYAIVLPESVQBIG--RSAPFRONGANLIFMGSKVKTGLGEMAFLSNR 580
DB      1546 ITEILNNNTTNE-----ERKNDIGPVRAAY-EGLNNI-----NAATTTG----- 1585

QY      581 LSHLDLSEKQLTEIPVQAFPSNALKEVLLPASLKTIREEAF-KKNHLKQL----- 630
DB      1586 -----DVTAKOTAVKQVQQLHANPVK---PAGKELDQAAADKKQTQETQTNASQQEIN 1638

QY      631 -----EVASALSHAFNALDDNGDEQDNKV-----VVKT--HHNSYALADGEHF-- 674
DB      1639 DAKQEVDTLNAQTN-VQSSSTNEYVDNAVKEGRKAKINAVKTFSEYKKDALAKIEDAYN 1697
```

```
QY      675 --IVDPD---KLSSTIVDL- EKILKLEGLDYSLRQTTQ-----TQPRDMT- 715
DB      1698 AKVNEADNSNASTSETAEAKQKLAELKQTADQNVNQATSKDDIEVQIHNDLDNINDYTI 1757

QY      716 TAGKA-----LLSKSNLRQEKQKFL---QEAQPFPLGRV-----D 747
DB      1758 PTGKESATTTDLAYADQKKNNISADTNATODEKQALKQVDQNVQTALESINNGVDNGD 1817

QY      748 LDKAIAKAKALVTKKATNGQOLLERSINKAVLAYNNSAIKKANVKRLEKELDLTG--- 804
DB      1818 VDDALTQGKAAI---DAIQVDATVKPKANQAI-----EVKAEDTKESIDQSDQLTAEK 1868

QY      805 -----LVEGKGPLAQATWQGVYLLKTPPLPEYVIGL---NVYFDKSGK--- 846
DB      1869 TEALAMIKQITDQAKQGITDATTTAEVEKAKAQ-----GLEAFDNIQIDSTEKQKAI 1920

QY      847 --LIVALDM-----SDTIGEGQKDAYGNPILNVDEDEGNYHALAVATLADY- EGLDI 895
DB      1921 EELETALDQIEAGVNVNADATTE-EKEAPTNAL-----EDILSKATEDISDQTTNAEI 1972

QY      896 KTIILNSKLSQTSITRQVPTAAHYRAGI-----FOAIQNAABAE----- 934
DB      1973 ATVKNSALEQLKQAKRINPEVKKNALREIVNVNQIIEIKNADADASAKETARTDLGRYF 2032

QY      935 ---QLLPKPGTHSEKS-----SSSESANSKDRGI-QSNPKTNRGHSAIL 975
DB      2033 DRFADKDKTQTNAEVAELOQNVTPITAEIVPQNDPDANDTNGDINDNDAFANSNANATP 2092

QY      976 PRTG 979
DB      2093 ENTG 2096
```

```
RESULT 74
US-10-080-334-234
; Sequence 234, Application US/10080334
; Publication No. US20040002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shimkets, Richard A
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Esha A
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zerhusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; TITLE OF INVENTION: Using the Same
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
```

PRIOR APPLICATION NUMBER: 60/278,796  
PRIOR FILING DATE: 2001-03-26  
PRIOR APPLICATION NUMBER: 60/281,521  
PRIOR FILING DATE: 2001-04-04  
PRIOR APPLICATION NUMBER: 60/276,677  
PRIOR FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: 60/311,595  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: 60/270,220  
PRIOR FILING DATE: 2001-02-21  
PRIOR APPLICATION NUMBER: 60/274,295  
PRIOR FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: 60/318,526  
PRIOR FILING DATE: 2001-09-10  
PRIOR APPLICATION NUMBER: 60/286,548  
PRIOR FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/291,765  
PRIOR FILING DATE: 2001-05-17  
PRIOR APPLICATION NUMBER: 60/270,797  
PRIOR FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: 60/276,400  
PRIOR FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: 60/270,810  
PRIOR FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 388  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 234  
LENGTH: 1348  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-10-080-334-234

Query Match 3.7%; Score 187; DB 15; Length 1348;  
Best Local Similarity 18.6%; Pred. No. 0.018;  
Matches 219; Conservative 159; Mismatches 344; Indels 454; Gaps 55;

QY 92 RQELKAVTENTSEKQITSGSQESLSLN-KTVPSTSNWEICDITTKGNTLVGLS 150  
DB 11 RQRELAQRSPKQKTAKLPSSQATSKCNLQAAPKTET-----ENTF-GQL 60

QY 151 KSGVEKLSGTDHLV--LPSQAADGTOLIQVAFAPDPPKKTAAEYTSRAGEN----- 201  
DB 61 KLTIBELDSLNNLRIRRIPEKAFDGLK-----DSLNELRANLLGDNLPST 109

QY 202 -----GEISQDLDVQKEI--INEGEVFNYSLLKKVTIPTGYKHIGQDAFVNKIAEY- 252  
DB 110 AELHVLKNLRLLDLSGNKIKLIEG-----LLK-----GCMDL-KEYIIDRNSLTSP 156

QY 253 ---NLPESELETISDYAFALHAKQIDLPDLNKAIGELAPFQNOITGKLSLPRQLMLAE 308  
DB 157 TNSLNGPSALR-----HLSLRO-----NQI-GSL-----LAD 182

QY 309 RAFKSNHIKTIIFRGNLSKVIGEASFQDNDLSQLMPLDGLKIESEAFNGPDDHYNNR 368  
DB 183 SFNAQRQLEIIDLNRHVRISDLSAPK-----GLQKIREIKLAG-----NR 223

QY 369 VVLWTKSGKVPGLATENTYVNDK-SLWQESPEDYTKWLEEDTYQKNSTGFSNKL 427  
DB 224 I-----SHLNSDVFELQSLQKLD-----LSENFFGQPTVALAAPVL 262

QY 428 QKVKRKNL--BIPKOHNGVTITEIGDNAFRVDFQNKTRKYDLEEVKLPSTIRKIGAF 485  
DB 263 KHNLSNNMLQDLYTHMQVV-----RSLESUDISRNTITI-----TPGTFRMGAL 310

QY 486 APDSNNLSFEASDDLEIEKEGAFMNRLETLELKD--LVTIGDAAFHINHIYVILPE 543  
DB 311 KYLDLSLSNLSRTIED--DALEGL---DSLQTLIIRKDNILLVPGSALGRLPOLTSIQLDY 365

QY 544 S-----VQETGRSAFRO-NGANNILFWGSKVKTIGEMA 575  
DB 366 NRVAALSABEILSGLOAGDITLTSRNVIRELPPGSGFQMFSSLSLHTLDLSGNSLAVINADT 425

QY 576 P--LSNRLEHLDLSEQK-----QLTEIPQAFSDNALKEVLLPAS-----LKTIRE 620

DB 426 FAGLESTLWALKLSQNRLTGLGGAPWLPPELSRLDSGNTLTE--LPSTIFEELENVQSL 483  
QY 621 AFKQVHLKQL-----EVASALSHIAFNALDNDGDQPD 654  
DB 484 NLSGNHLTFLTGALFKPLDLRLQVIDLSCNIRQISGDLIAGLQDLKHIYINDNQLEQD 543  
QY 655 NKVV-----VKTHNSYALADGHEF----- 674  
DB 544 GSFVNLWNISSIDLNNRIGSIRSAFVNVWMLQKDLHGNQLSAFKGYFTGTGIBEL 603  
QY 675 -IVD-----PDKLSSTIVDLEKI-----LKLIEG 697  
DB 604 DISDNQLSVLPSSPRIHPRLREIRAANNKFSFFPAELISTIQLEHIDLSHNLKTIEE 663  
QY 698 LDYSTL-----RQTTQTQFRDM-----TTAGKALLSKS 725  
DB 664 LDFARLPRLLVANNQLDMVSEMAFHNSTQLDILAHNNLDRIGERTFEGVRLLEQL 723  
QY 726 NLRQG-----EKQKF-----LQEAQOFFELGRVLDL---KA 751  
DB 724 NLEGNRLSELSDGVFERTKQMLENINLAHNRFEYAPLNALQRFVSSVDLSHNKIKE 783  
QY 752 IAKAEKALVTKATKNGQLLERSINKAVLAYNNSAIKKA---NVKRLKELDLTLGLVE 807  
DB 784 LPGDDSIWV-----NIKRIDLSPNPLSSKAVHNVLEPKTVRELSL----- 824  
QY 808 KKGPLAQATMVGQVYLLTPTPLPBYIYGLNVYFDKSGK-----LIIVALDMSDTI 857  
DB 825 -----ACTGIENLELETP-----FLOFLNLSHNKLKNVKEPVFQVTLTLETLSSNQ 873  
QY 858 GEGOKD-AYGNPILNV---DEDNEGYHALAVATLADYEGLDIKTILNSKLSQLSI--- 909  
DB 874 LESUEDLSNAMPQLQVLOSLDVSNNFSFIVSQNSFGKLEML--RSLRLSHLPQCTRIEKN 931  
QY 910 --RQVPTAAHYRA-----GIFQ-----AIONAAAEQLLP----- 938  
DB 932 AFKQPLNVLSEAYDPLPILGYLDLQSLILELPGLEVLQIEVKDSSIGSEQIQPLKHPRLK 991  
QY 939 KPGTHSEKSSSESANSKDRGLQSNPKTNRGHSAI 974  
DB 992 SLGIRGDRLKSTSSGTL--AGLKSNDLSVQLRNTSL 1025

RESULT 75  
US-10-369-493-1950  
; Sequence 1950, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xiaofeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 1950  
; LENGTH: 1312  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-1950

Query Match 3.7%; Score 186.5; DB 15; Length 1312;  
Best Local Similarity 19.0%; Pred. No. 0.018;  
Matches 200; Conservative 174; Mismatches 362; Indels 315; Gaps 49;

QY 21 NQEVPSLVKEPILKOTQASSISGADYAESGSKLKINETSGPVDDTVDLFSDKRTPP 80



Db 800 -TOSTSKESLSNELTELKIAAKK---EQACKGEEDNLARLKKELTETELAKKEADLS 855  
QY 638 HIAFNALDNDGDEQFNDKVVVTHNSVALADGEHFIVDPKLSSTI--VDLEKILKLE 696  
Db 856 PLTSEMSSTSEKLEEA-----AKKNDKTKTIELIALRRDQRIKLOH 901  
QY 697 GLDYSTLRTOQTQFDMTTAGKALLSKSNLRQGEKQKFLQBAQFFLGR--VDLOKAI-- 752  
Db 902 GLD-----TYERELKEMRLYK-----QKTLTKDBEVKGRMEVELDNLLQY 944  
QY 753 -----AKAEKALVT--KKATNGOLLERSINKAVLAYNNSAIKANKVKLEKEL 799  
Db 945 LBEVSLSEGAKEQYLETDPPEARVRVKKLJKAIEEL-----GTVNLGSIDEFERNV 998  
QY 800 DILLTGLVEGKGLAQA--TWVQCVLLKTPLP--LPEYVIGLVVYFDKSGKLIYA----- 850  
Db 999 ERYKFLSEKEDLTEAKNTLFQVIEEMDEMYKRFNDTFQVIRSHPDQVFRSLFGGGR 1058  
QY 851 ----LMSDTIGGQK---DAYGNPILNVDENEGYHALAVATLADYEGLDIKTILNSKLS 904  
Db 1059 LRLTPNDLLHSGVEIIAQPFGKQLNLLSGGERALTAL-----LF 1103  
QY 905 QLTSTROVP-----TAAYHRAGIFQAIQ 927  
Db 1104 SILKVRPVPFCVLDEVEAALDEANVFRFAQ 1133

## RESULT 77

US-10-282-122A-46664  
; Sequence 46664, Application US/10282122A  
; Publication NO. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA 034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 46664  
; LENGTH: 1788

; TYPE: PRT  
; ORGANISM: Bacillus anthracis  
US-10-282-122A-46664  
Query Match 3.7%; Score 185.5; DB 15; Length 1788;  
Best Local Similarity 19.2%; Pred. No. 0.033;  
Matches 223; Conservative 158; Mismatches 397; Indels 383; Gaps 57;  
QY 2 KKHAKTVALTITTVSVVTHNQEVSLVKEPILKQTAQSSISGADYAESSGSKLKINET 61  
Db 582 QNNIKTDALLKDIPTDTHELDVSIV-----VKNASYDD--KRLVTGDT 625  
QY 62 SGPVDVTVT-----DL-FSD-----KRTTPEKIKDNLAGPREQELKAVTENT 103  
Db 656 VN--NYTVTNKNGFDLQFNEDINSAYVITYKTPNNVIKDGKIK-----NTVTADN 676  
QY 104 ESEKQITSGSQLEQSKESLSLNT--VFSTSNWEICDFITKGTNLVGLSKSV----- 154  
Db 677 GSSKENEAGFOOQNIKSNNAETNYKDKTTTWT-----ITVNNNNYPLNNAIITDFDHG 732  
QY 155 -----EKL-----SOTDHLVLPQAAADGTOLIQVASFAFPPDKKTAIAEYTS-- 196  
Db 733 GLQLKDKKLEIKDGYDTLOAETDYLVDVTDKGFKITLIGTYOSNMT---KTLIVKITDFF 789  
QY 197 -----RAGE-----NGEISQLDVGKBEINEGEV--FNSYLLKKVKTPTGYKHIGODAFVD 245  
Db 790 DYTLESCKTSFKNTGNLSWIDAGSNPQSNKVEADFPDPTFKA---NGYKYGSYNAQTK 846  
QY 246 -----NKNIAEVNLPESLETISD-----YAFAHALAKQIDLDPNLKAIGELAPFF 289  
Db 847 EITWIIIGFNNYNNVEIKDPYVIDVIQDKQKLVPGSIEVRHMLN--GSPDNARP-GD--- 899  
QY 290 DNQITGKLSLPRQLMLAERAPKSNHKTIFRG-----323  
Db 900 -----AVPIEQYELEPTDKNKTLOVHFQKOSINSPYIIFKTSLDGELIQNTYKNE 951  
QY 324 -----NSLK-----VIGEASFQDNDS--QLMLPDGLKEIESEAFNGPG 361  
Db 952 AELKDGSKI VNTLKGDAQVNGKGSFVTKKAVQDDNYINNSAINESQSTIADAVVDDPT 1011  
QY 362 D-----DHYNNRVVLWTKSGKNPSGLATE--NTYVNPDKSLWQESPEIDY 404  
Db 1012 DNQVLVEDSFHLYPTTVDYGN-----VTKDTANELKQGTDYKLTITTDNNTGKHFEIAF 1067  
QY 405 TKWLEEDFTYQKNSVTG-----FSNK-----GLQKVKENK-----434  
Db 1068 LKKIDRAVILRYRSLINADDKEKVSNAKAKIAGNQTVTKNTETVETIEVRMSSGGSGSAT 1127  
QY 435 -----NLEIPKQHN-----GVTTIT---EIGDNAFRNV-----DFQNKTLRKYDLE 471  
Db 1128 NGRGNLEIIVKVDNDNKVPLSGAEFTLYDRGTGTVIRKITTDDKQIAKFNLLKRDYLLK 1187  
QY 472 EVKLPTSTIRKIGAFAPQSNNLKSPASDDLEIEKEGAPMNNRIETLELKDVLVTGDAAF 531  
Db 1188 ETKAP-----EGYVISWDLKQKIVELGSOBTITYKLANK-----1222  
QY 532 HINHIYAIVLPESVQEIERSAFRQNGANNLIFMGSKVKTGLGEMAFSLNRLEHLDSEKQ 591  
Db 1223 ---KFVGKVLLTKS--DDLKNVTLOQAVFTLDDKKKI-----ISEHEKLTNDQOQ 1270  
QY 592 LTEIPVQAFSDNALKEVLLPASL---KTIREEAFKKNHLKQEVASALSHIAFNALDND 648  
Db 1271 IT-----VDN-----LKPGTYVLOETTAPEHYKIDS-----TPIQTIKED-- 1306  
QY 649 GDEQFDNKVVVTKTHNSYALADGBHFIVDPDKLSSTIVDLEKILK-----LIEGLDYSTL 703  
Db 1307 -----QTTVINRATNS--LIPGSAILTKVDKDKGTLAGAEFSVRDRHNNVIRG----- 1353  
QY 704 RQTTQTRQDMTTAGKALLSKSNLRQGEKQKFLQE---AQFPLGRVLDLKAIAKAEKALV 760  
Db 1354 -----YEKUTTNDQOQIEATNLRPGDYQ--FVEEKAPKQVDIDIKTIEFTIVKSQKAV 1405  
QY 761 TKKATK-----NGOLLERSINKAVLAYNNSAIKANKVVRLEKELDLTLGL 805

Db 1406 TVTATNHLKGGVTLTKTDDIDGTALAGAFKIVDANDEKKVIRENVK-----TG- 1455  
QY 806 VEGKPLAQATWQVGY-LLKPLPLPYIYGLN---VYFKSGKLIYALDMSDITGEGQ 861  
Db 1456 --ADGKVTWQLEPGTYKFIETAP-KDYVLNANPIEFITDKSQSFATVATNLSLKTGE 1512  
QY 862 KDAGNPILNVD--DNEGHALAVATLADYEGLDIKILNSKLSQLTSIRQVPTAAVHR 919  
Db 1513 VE-----LLKVDGFGDKKPLKG-AVFVKIVDNNNDVTRDLTDTADGKTAKDKL-----R 1560  
QY 920 AGIFOAIONAAAEQQLPKP 940  
Db 1561 PGTYKFIETAPVHLRAEP 1581

## RESULT 78

US-09-815-242-5639  
; Sequence 5639, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5639  
; LENGTH: 2086  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-5639

Query Match 3.6%; Score 185; DB 9; Length 2086;  
Best Local Similarity 18.3%; Pred. No. 0.043;  
Matches 201; Conservative 150; Mismatches 417; Indels 328; Gaps 41;  
QY 14 TVSVYTHNQEVFSLVKEPILKQTOASSISGADYAESGSKLKI-NETSGPVDVTDL 72  
Db 595 TPVTVTTSSEV-----TAANQVNSAK-QELNGDERLVRKQNAITADLTQL 642  
QY 73 FSDKRTPPE--KIKDLNKGPREQELKAVTENTESEKQITSGSQLEQSKESLSLNKTVPS 130  
Db 643 -----NTPQAKLKQGVQGNLELDIQTVTQNGALNNAMKG-----LRDSIANETTVKA 692  
QY 131 TSNWEIC-----DPIKGNLTVLGSKGV----- 154  
Db 693 SQNYTDASPNNOSTYNSAVNAKGIINGTNNPMTDTSAITQATQVNNAKNGLGAENLR 752  
QY 155 -----EKLSTQDHLVLPQAAADGTQLIQVASFAFTPKDKTAIAEYTSRACENGESIQ 207

Db 753 NAQTAQNLTLSHL-----TNQKSAISSQIDRAGHVSEVTA 792  
QY 208 DVDGKEIINEGEVFNYSVLLKVVITPTGYKHIGODAFVNDKNIAEVLNLPESLETISDYAPA 267  
Db 793 KNAATLNTMGNEQAIDHQNTQVKGWFTDAD---KAKRDAVTNAVSRAETILNKTOG 849  
QY 268 HLAUKQDIDLPNLKAIAGELAFFDQITGKLSLPQMLRABAFKSNHIKTIERFGSLK 327  
Db 850 ANTSKQ-----DVEAAIQNVTSKALNGDQNV-----TNAKNTAKHALNNLT 892  
QY 328 VIGBASPOD--NDLSQLMPLDGLBKIESEAFTPGDDHYNNRVVLTWKSKNPSGLATE 385  
Db 893 SINNAQKRDLTTKIDQATTAGVAVSTGTQLTMTANLQNGI-----NDKANTLASE 946  
QY 386 NTYVNPDKSLMQESPEIDYTKWLEDEFTYQKNSVTGFSNKGLOKQYKRNKNLEIPKOHNGV 445  
Db 947 N-YHDADSD-----KKTAYT-----QAVTNAENI----- 969  
QY 446 TITIGDNAFRNVDPQNTLKYDLEVKLPSTTRKIGAPAFQSNNLKSFASDDLEIK 505  
Db 970 ----LNKNSGSNLD---KAAVENALSQV-----TNAKGALNGHNLEQAK 1007  
QY 506 EGAFMNNRIETLE-----LKDKLVTIGDAAPHINHIIVLPESVQEIGRSAFRQNGANN 560  
Db 1008 SNA--NTTINGLOHLTTAOKDKLQOQVQAQNV-----AGVDTVKSSANTLNGA-- 1054  
QY 561 LIFMGSKVTLGEMAFILSNRLEHLDLSEKQKLTETIPVQAFSDNALKEVLLPASLKTIRBE 620  
Db 1055 ---MGLTRNSIQDNTATNNGQNYLDATESN----- 1081  
QY 621 AFKQNLKQLEVASALSIAFNALDDNDGDQPNKVVVKTTHNSYALADGEHFIIVDPDK 680  
Db 1082 --KTYNNAVDSANGV-----INATSNPNMDANAIQATQVTTSTKNAL-DGTHNLTOAKQ 1134  
QY 681 LSSTIVD----LEKILK-----LIEGLD----- 699  
Db 1135 TATNAIDGATNLNKAQDKALKAQVTSQORVANVTTSIQOTANELNTAMGOLQHGIDDENAT 1194  
QY 700 ----YSTLRQTOTQFRDMTTACKALLSK---SNLRQGEKQKFLQEA---QFFLGRVD 747  
Db 1195 KQTKYRDAEQSKKTAYDQAVAAAKAILNKQTSNSDKAAVDRAALQQVTSSTKDALNGDAK 1254  
QY 748 LDKAIAKAEKALVTYKATNGQ---LLERSINKAVLAYNNSAIKANVVKRLEKELDLLTGL 805  
Db 1255 LAEAKAAAKQNLGTLNHTNAQRTALEGQINQOATTVDGVNTV-KTNANTLDGAMNSLQGS 1313  
QY 806 VEGKPLAQATWQVGYLLKTPPLPEYVIGLVVYFDKSGKLIYALDMSDITGEGQDAY 865  
Db 1314 INDK-----DATLRNQNYL-----DADESKKNAYTQAVTAAEGILNKQTG 1353  
QY 866 GNPILNVDENEGYHALAVATLADYEGLDIKILNSKLSQLTSIRQVPTAAVHRAGIFQA 925  
Db 1354 GN-TSKADVDN-----ALNTVTTRAKAALNGAENLNTKTSATNTINGLPLNLQLOK----- 1403  
QY 926 IONAAAABEQLLPKPGTH--SEKSSSESANSGKQGLQSNPKTNRGRHSAILPRTGSKGS 983  
Db 1404 -DNLKHQVEQANVAGVNGVKDKGTGLNTAMGALRTSIQNDNTTKTSQNYLDASDSNKN 1462  
QY 984 F-----VYGILGVTS 993  
Db 1463 YNTAVNNANGVINVTN 1478

## RESULT 79

US-10-080-608A-11  
; Sequence 11, Application US/10080608A  
; Publication No. US20030198956A1  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; APPLICANT: Williams, Mark  
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES





Db	459	QMEEMKTRHKGEMENALSYSNITVNEQI KLMNVAINELNI KLQDNTNSOKELKBEGL	518
Qy	407	WLEEDFTYQKN-----SVTGSNKGLOKVKR---NKNLEIPKOHNGVTITE-----IGD	452
Db	519	ILIEKALORQLEDVLBELSFSRBOIQAROTIAEQESKNEAHKSLSTVEDLKABIVSA	578
Qy	453	NAPR-----NVDFQNKTRKYDLEEVKLPSTIRKIGAFAPOSNNLKS---FEASDDLE	502
Db	579	SSSRKEULEKHEAVTNVYKILMLEKEKNVLDORMAESQEAELERLRTQLLFSHEELS	638
Qy	503	EKEGAFPMNRIETLELKOKLVITGDAAFHNIHIAVLPEVQETGRSAPRONG-----A	558
Db	639	KLKEDLEIHRINIEKJDNL-----GIHYKQOIDGLQNMESOKIETMFOEKDNLITQ	692
Qy	559	NNLI FMGSKVKTGLBMAFLNRLHLDLSEQOLTEIPVQAFSDNALKEVLLPASLKTIR	618
Db	693	NQLILEISKLOL-QOSLVNSKSEMTL-----QINE-----LQKEIEIR	732
Qy	619	BEAPKQVHLKQ-LEVASALSHIAFNALDDNDGDSQDPNKVVVKTTHNSYALADGEHF	677
Db	733	QOEKEKGTLEQEVQBLQLKTELLEKQMKKEKENDLQ-----EKPAQLEAENSILK	781
Qy	678	PDKLSSTIVDLKILKLIEGLDYSLTRQTTQTFQDMWTAGKALLSKSNLRQEKQKFLQ	737
Db	782	DEKKT-----LEDMLKI-----HTVPVQERELIFLDSIKS-----KSKOSVWEKEI	827
Qy	738	EAQFFLGR-VDLDDKAIAKAERKALVTKATKNGOL-----LERSINKA	778
Db	828	ENEDLKQOQIQLNIEIEKQBNTP--SPAEKKFVFNVYQBELQEVACLLKVDDLEDSK	885
Qy	779	VLAYNNSAIKXANVXRLEKELD	801
Db	886	ELEY-----KSKLALNEELHL	902

Qy	453	NAPR-----NVDFOKTLRKYVDLBEVVKLPSTIRKIGAPAFQSNILKS-----FEASDDLE	502
Db	579	SESRKELELGHAEAVTNKYIKLEMLEKEKNVLDMAESQEAELERLTQTLFLFSHEELS	638
Qy	503	EIKEGAPMNNRIETLELKDKIATVIGDAAFHINHIYAIVLPESVQEIGRSAPQNG-----A	558
Db	639	KLKEDLEIHRINIEKLNKL-----GIHYKQDGLQNEWSQKIETHQFEKDNLIITKQ	692
Qy	559	NNLIFMGSKVKTIGEMAFLSNRLHLDLSEQQLTEIPVQAFSDNALKEVLLPASLKTR	618
Db	693	NQLILEISKLKDL-QQSLVNSKSEMTL---QINE-----LQKEIBILR	732
Qy	619	ERAFKQNHUKQ-LEVASALSHIAFNALDNDGDGEQPDNKKVVVTHHNSVALADGHEFVD	677
Db	733	QBEKEKGTLEQVEQLQKTELLEKQMEKENDLQ-----EKPAQLEAENSILK	781
Qy	678	PKLSSTIVDLEKILAKLIEGLDYSLRQTQTQFPDMTTAGKALLSKSNLRQGEKQKFLQ	737
Db	782	DEKKT-----LEDMLKI-----HTPVSQEBRIFLDSIKS-----XSKDSVWSEKIEILIE	827
Qy	738	EAQFFLGR-VDLLDKAIKAKEKALVTKKATKNGOL-----LERSINKA	778
Db	828	ENEDLKQOQIQLNEEIEKQBNTF--SFAEKNFVNVYQELQEBEVACILKVKYDLDLSDKNKQ	885
Qy	779	VLAYNNSAIKANVVKLEKELD	801
Db	886	ELEY-----KSKLKALNEELHL	902

RESULT 82  
US-10-756-149-5739  
; Sequence 5739, Application US/10756149  
; Publication No. US2005018175A1  
; GENERAL INFORMATION:  
; APPLICANT: Ariz, Natasha  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: NOVEL METHODS OF

```
/ TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
/ FILE REFERENCE: file
/ CURRENT APPLICATION NUMBER: US/10/756,149
/ CURRENT FILING DATE: 2004-01-12
/ NUMBER OF SEQ ID NOS: 5818
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 5739
/ LENGTH: 3908
/ TYPE: PRT
/ ORGANISM: Homo Sapiens
US-10-756-149-5739

Query Match          3.6%; Score 185; DB 18; Length 3908;
Best Local Similarity 19.5%; Pred. No. 0.1;
Matches 180; Conservative 164; Mismatches 349; Indels 230; Gaps 44;

QY 7 TVALTLTVSVTHNQEVPSLVKEPILKOT--QASSISGADYAESGSKSLKINETSGP 64
Db 82 TIMRTLHSGEITSHEQG--FSVELESEISTTADDCSEVNGCSFVMRTGKPTNLLREBEFG 140
QY 65 VDDTVTDL-FSDKRTTPEKIKONLAKPREQELKAVTENSEKEKQITGSGQ-LEOSKESL 122
Db 141 VDDSYEQAGQADSPHLEWMESELA--GKQHEIEELNRELE-EMRVYTGTEGLQLOEPE 197
QY 123 SLNK-----TVPSTSNWEICDPFITKGNLTGLVGSKGVEKLSQTDHLVLPQADGTLQ 177
Db 198 AAIKQORDGIITQLTANLQAR-REKDETM----REFLELTEQSKLQIQFQQLQASETLR 252
QY 178 VASFAFTP-----DKTAAIAYTSRAGENGESIQOLDVDGKEIINEGEVFN---SYLLKKVT 230
Db 253 NSTHSSTAADLLQAKQOQILTHQOOLEEQDHLE-DYQKKK-----EDFTMQISFLOEKIK 306
QY 231 IPTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAFAPHLAKQIDLPDNL-----KAIGEL 286
Db 307 V-----YEMEQDKKVENSKKEELOEKETIIEELNTKIIIEBKTLLEKOKLTADKLGL 362
QY 287 AFFDNOITG-----KLSLP-----RQLMLRAERAFKSNHIKTIIFRNGS 325
Db 363 ---QEIQVQKNOEIQNMKLELTNSKQKERSSEBIKQLMGTVEELQKRNH-KDSQF---- 414
QY 326 LKVIQGEASFQDNDLSQMLPD---GLEKIESEAFNGPGDDHNNRVVLWTK----- 374
Db 415 -----ETDIVORMEQETORLEQRAEL-----DEMYGQIQVQMKOELIRQHMA 458
QY 375 -----SGKNPGLAT-ENTYVNPDK-----SLWQSPESDIYTK 406
Db 459 QMEEMKTRKHGEMENALRSYNTVNEQIKLMNVAINELNIKQDTNSQKSLKEELGL 518
QY 407 WLEEDFTYQKN-----SVTGFSGNKGLOKVKR---NKNLEIPKQHNGVTITE-----IGD 452
Db 519 ILBEKCALQROLEDLVEELSFSREQIQARQTIAEQESKLEAHKSLSTVEDLKAEIVSA 578
QY 453 NAFR-----NVDQFONKTLKYDLBEVKLPSTIRKIGAPAFQSNLKS---FEASDDLE 502
Db 579 SESRKELELKHEAEVNTYKIKLEMLEKNAVLDRMAESQAELELRLOLFLFSHEBELS 638
QY 503 EIKEGAFMNRITELKDKLVITGDAAAPHINHIYAIVLPESVQEIQRSAFRQNG-----A 558
Db 639 KLKBDLEIHRINIEKIKLKNL-----GIHYKQIQIDGLQNMESQKTIQTFEKNLITKQ 692
QY 559 NNLIIFMGSKVKTIGEMAFNSRLNLEHLDLSEQKQLEIPVQAFSDNALKEVLLPASLKTIR 618
Db 693 NQILILEISKLOL-QQSIVNSKSEMTL-----QINE-----LQKEIETLR 732
QY 619 ERAFKGNHLKQ-LEVASALSHIAFNALDNDGDEQFDNKKVVKTHNSVALADGEHFIVD 677
Db 733 QBEKEKGTLEQEVQELQKTELLEKQWKEKENDLQ-----EKAQLEAENSILK 781
QY 678 PKLSTVIVLEKILKILIEGLDYSTLRQTTQTFRDMTTAGKALLSKSLNRQGEQKFLQ 737
Db 782 DEKKT-----LEDMLKI-----HTPVSQBERLIFLDSIKS-----KSKDSVMEKEIEILIE 827
QY 738 EAQFFLGR-VOLDKATAKAEKALVTKKATKNGQL-----LERSINKA 778
```

```
Db 828 ENEDLKQOCIQLNIEIERQNTF--SFAEKNFVNYQELQEBYACLLKVKDDLEDKKNQ 885
QY 779 VLAYNNSAIKKANVKRLEKELD 801
Db 886 ELEY-----KSKLKALNEELHL 902

RESULT 83
US-10-370-685-100
/ Sequence 100, Application US/10370685
/ Publication No. US20030215903A1
/ GENERAL INFORMATION:
/ APPLICANT: Hyman, Paul
/ TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
/ FILE REFERENCE: NANF.P-004
/ CURRENT APPLICATION NUMBER: US/10/370,685
/ CURRENT FILING DATE: 2003-02-21
/ PRIOR APPLICATION NUMBER: 10/080,608
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 159
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 100
/ LENGTH: 3911
/ TYPE: PRT
/ ORGANISM: human
US-10-370-685-100

Query Match          3.6%; Score 185; DB 15; Length 3911;
Best Local Similarity 19.5%; Pred. No. 0.1;
Matches 180; Conservative 164; Mismatches 349; Indels 230; Gaps 44;

QY 7 TVALTLTVSVTHNQEVPSLVKEPILKOT--QASSISGADYAESGSKSLKINETSGP 64
Db 94 TIMRTLHSGEITSHEQG--FSVELESEISTTADDCSEVNGCSFVMRTGKPTNLLREBEFG 152
QY 65 VDDTVTDL-FSDKRTTPEKIKONLAKPREQELKAVTENSEKEKQITGSGQ-LEOSKESL 122
Db 153 VDDSYEQAGQADSPHLEWMESELA--GKQHEIEELNRELE-EMRVYTGTEGLQLOEPE 209
QY 123 SLNK-----TVPSTSNWEICDPFITKGNLTGLVGSKGVEKLSQTDHLVLPQADGTLQ 177
Db 210 AAIKQORDGIITQLTANLQAR-REKDETM----REFLELTEQSKLQIQFQQLQASETLR 264
QY 178 VASFAFTP-----DKTAAIAYTSRAGENGESIQOLDVDGKEIINEGEVFN---SYLLKKVT 230
Db 265 NSTHSSTAADLLQAKQOQILTHQOOLEEQDHLE-DYQKKK-----EDFTMQISFLOEKIK 318
QY 231 IPTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAFAPHLAKQIDLPDNL-----KAIGEL 286
Db 319 V-----YEMEQDKKVENSKKEELOEKETIIEELNTKIIIEBKTLLEKOKLTADKLGL 374
QY 287 AFFDNOITG-----KLSLP-----RQLMLRAERAFKSNHIKTIIFRNGS 325
Db 375 ---QEIQVQKNOEIQNMKLELTNSKQKERSSEBIKQLMGTVEELQKRNH-KDSQF---- 426
QY 326 LKVIQGEASFQDNDLSQMLPD---GLEKIESEAFNGPGDDHNNRVVLWTK----- 374
Db 427 -----ETDIVORMEQETORLEQRAEL-----DEMYGQIQVQMKOELIRQHMA 470
QY 375 -----SGKNPGLAT-ENTYVNPDK-----SLWQSPESDIYTK 406
Db 471 QMEEMKTRKHGEMENALRSYNTVNEQIKLMNVAINELNIKQDTNSQKSLKEELGL 530
QY 407 WLEEDFTYQKN-----SVTGFSGNKGLOKVKR---NKNLEIPKQHNGVTITE-----IGD 452
Db 531 ILBEKCALQROLEDLVEELSFSREQIQARQTIAEQESKLEAHKSLSTVEDLKAEIVSA 590
QY 453 NAFR-----NVDQFONKTLKYDLBEVKLPSTIRKIGAPAFQSNLKS---FEASDDLE 502
Db 591 SESRKELELKHEAEVNTYKIKLEMLEKNAVLDRMAESQAELELRLOLFLFSHEBELS 650
```

QY 503 EIKGAFMNNRIETLEKDLVLTIGDAAPHINHIYAIVLPESVQIEGRSAFRQNG----A 558  
 Db 651 KLKEDLEIHRINIEKLKDNL-----GIHYKQIDGLQNEMSQKIETMQEKNLITKQ 704  
 QY 559 NNLIIPMGSKVKTGLGMAFLSNRLEHLDLSEKQQLTEIPVQAFSDNALKEVLLPASLKTIR 618  
 Db 705 NQLILEISKDL-QQSLVNSKSEMTL-----QINE-----LQKEIEILR 744  
 QY 619 EBAFKNNHLKQ-LEVASALSHIAPNALDDNDGDEQFDMKVVVKTTHNSYALADGHEFIVD 677  
 Db 745 QBEKEGTLEQVQELQNLTELLERQMKENDLQ-----EKPAQLEAENSILK 793  
 QY 678 PDKLSSTIVDLKILKLEGLDYSLRTQTQTPRDMTTAGKALLSKNLRQGEKQKPLQ 737  
 Db 794 DEKKT-----LEDMLKI-----HTPVQERLIPLDSIKS-----KSDSVWEKEIEILIE 839  
 QY 738 EAQFFLGR-VLDLKAIAKAELVTKKATKNGQL-----LERSINKA 778  
 Db 840 ENEDLKQCCIQLNEIEKQRNTF--SFAEKNFVNYQELQEBYACLLKVKDDLEDKKNKQ 897  
 QY 779 VLAYNNSAIKXANVRLSEKELD 801  
 Db 898 ELEY-----KSKLKALEELHL 914

## RESULT 84

US-10-408-765A-1839  
 ; Sequence 1839, Application US/10408765A  
 ; Publication No. US20040101874A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ghosh, Soumitra S.  
 ; APPLICANT: Fahy, Boin D.  
 ; APPLICANT: Zhang, Bing  
 ; APPLICANT: Gibson, Bradford W.  
 ; APPLICANT: Taylor, Steven W.  
 ; APPLICANT: Glenn, Gary M.  
 ; APPLICANT: Warnock, Dale E.  
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
 ; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
 ; FILE REFERENCE: 660088.465  
 ; CURRENT APPLICATION NUMBER: US/10/408,765A  
 ; CURRENT FILING DATE: 2003-04-04  
 ; NUMBER OF SEQ ID NOS: 3077  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1839  
 ; LENGTH: 3911  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-408-765A-1839

Query Match 3.6%; Score 185; DB 16; Length 3911;  
 Best Local Similarity 19.5%; Pred. No. 0.1;  
 Matches 180; Conservative 164; Mismatches 349; Indels 230; Gaps 44;  
 QY 7 TVALITTVSVTVHNOEVPSELVKEPIKLT--QASSISGADYAESGSKLKNETSGP 64  
 Db 94 TIMRTLHSGEITSHEQG-FSVELESEISTADCSSEVNGCSFVMTGKPTNLLREBEFG 152  
 QY 65 VDDTVTL-FSDKRTTPEKIKDLAKPREQBLKAVTENTESEKQITSGSQ-LEQSKESL 122  
 Db 153 VDDSVSEGAQDSPTHELMWSELA--GKHIEHLNRELE-EMRVTVTEGLQQLQEF 209  
 QY 123 SLNK-----TVPSTGNWEICDFTKNTLVGLSKSGVEKLSQTDHLVLPSSQADGTQIL 177  
 Db 210 AAIKQDGIITQLTANLQAR-REKDETM-----REFLELQESQKLQIQFQQLQASLIR 264  
 QY 178 VASFAFTP-----DKKTAIAEYTSRAGENCEISQLVDGKEIINEGEFVN---SYLLKKVT 230  
 Db 265 NETHSSTAADLLQAKQQLIHTHOQLEEQDHILLE-DYQKKK-----EDFTMQISFLQEKIK 318  
 QY 231 IPTGVKHIQDQAVFQVKNIAEYVNPESLETISDYAFALHAKQIDLPNML-----KAIGEL 286  
 Db 319 V-----YEMQDKVNSNKEIEQKETIIEELNTKIEBEKTKLEKOKLTADKLGL 374

QY 287 APPDNLITG-----KLILP-----ROLMLABEAPKSNHIKTIETFRGNS 325  
 Db 375 ---QEQIVQKNQETKNMKLELTNSKQEROSSEIEIKQMGTVLELQKRNH-KDSQF----- 426  
 QY 326 LKVIGEASFQNDLSQLMLPD---GLEKISEAFTGNPDGDDHYNRRVVLWTK----- 374  
 Db 427 -----ETDIVQREQETQRLQLRAEL-----DEMYQQQIVQMKQLIRQHMA 470  
 QY 375 -----SGKNPSGLAT-ENTVYNPDK-----SLMQESPEIDYTK 406  
 Db 471 QMBEMKTRHKGEMENALRSYNTVNSDQIKLMMVAINELNKLQDTSQKEKLKEELGL 530  
 QY 407 WLEEDFYQKN-----SVTGFSGNGLQKVKR---NKNLEIPKONGVTITE-----IGD 452  
 Db 531 ILSEKCALQRLQEDLVEELSPSREIQARQTTIAEQSKLEAHKSLSTVEDLKAELVSA 590  
 QY 453 NAFR-----NVDFQNKTLRKVDLEVKLPSTIRKIGAFAPQSNNLKS---FEASDDLE 502  
 Db 591 SESRKELELKHAEAVTNYKIKLEMLEKNAVLDRMAESQEAELERLTQLLFSHEBELS 650  
 QY 503 EIKGAFMNNRIETLEKDLVLTIGDAAPHINHIYAIVLPESVQIEGRSAFRQNG----A 558  
 Db 651 KLKEDLEIHRINIEKLKDNL-----GIHYKQIDGLQNEMSQKIETMQEKNLITKQ 704  
 QY 559 NNLIIPMGSKVKTGLGMAFLSNRLEHLDLSEKQQLTEIPVQAFSDNALKEVLLPASLKTIR 618  
 Db 705 NQLILEISKDL-QQSLVNSKSEMTL-----QINE-----LQKEIEILR 744  
 QY 619 EBAFKNNHLKQ-LEVASALSHIAPNALDDNDGDEQFDMKVVVKTTHNSYALADGHEFIVD 677  
 Db 745 QBEKEGTLEQVQELQNLTELLERQMKENDLQ-----EKPAQLEAENSILK 793  
 QY 678 PDKLSSTIVDLKILKLEGLDYSLRTQTQTPRDMTTAGKALLSKNLRQGEKQKPLQ 737  
 Db 794 DEKKT-----LEDMLKI-----HTPVQERLIPLDSIKS-----KSDSVWEKEIEILIE 839  
 QY 738 EAQFFLGR-VLDLKAIAKAELVTKKATKNGQL-----LERSINKA 778  
 Db 840 ENEDLKQCCIQLNEIEKQRNTF--SFAEKNFVNYQELQEBYACLLKVKDDLEDKKNKQ 897  
 QY 779 VLAYNNSAIKXANVRLSEKELD 801  
 Db 898 ELEY-----KSKLKALEELHL 914

## RESULT 85

US-10-171-311-8  
 ; Sequence 8, Application US/10171311  
 ; Publication No. US20030087270A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schlegel, Robert  
 ; APPLICANT: Chen, Yan  
 ; APPLICANT: Zhao, Xumei  
 ; APPLICANT: Monahan, John  
 ; APPLICANT: Kamatkar, Shubhangi  
 ; APPLICANT: Giatt, Karen  
 ; APPLICANT: Gannavarapu, Manjula  
 ; APPLICANT: Hoarsh, Sebastian  
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
 ; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
 ; TITLE OF INVENTION: OF CERVICAL CANCER  
 ; FILE REFERENCE: MRI-035  
 ; CURRENT APPLICATION NUMBER: US/10/171.311  
 ; CURRENT FILING DATE: 2002-06-12  
 ; PRIOR APPLICATION NUMBER: US 60/298,159  
 ; PRIOR FILING DATE: 2001-06-13  
 ; PRIOR APPLICATION NUMBER: US 60/298,155  
 ; PRIOR FILING DATE: 2001-06-13  
 ; PRIOR APPLICATION NUMBER: US 60/335,936  
 ; PRIOR FILING DATE: 2001-11-14  
 ; NUMBER OF SEQ ID NOS: 238  
 ; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8  
; LENGTH: 3917  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-171-311-8

Query Match 3.6%; Score 185; DB 14; Length 3917;  
Best Local Similarity 19.5%; Pred. No. 0.11;  
Matches 180; Conservative 164; Mismatches 349; Indels 230; Gaps 44;

```
QY 7 TVALTLTVSVTHNQEVSLVKEPIKQT--QASSISGADYAESGSKLKINETS GP 64
DB 82 TMRTLHSGEITSHQG-FSVELESEISTTADDCSEVNGCSFVMRTGKPTNLLRREEFG 140
QY 65 VDDTVTDL-FSDKRTTPEKIDNLAQPREQELKAVTENSEKQITSGSO-LEQSKESL 122
DB 141 VDDSYSEQAQDSPTHELMWSELA--GKQHEIEELNRELE-EMRVYTGEGLOQLQEF 197
QY 123 SLNK-----TVPSTSNWEICDFTKGNLTVGLSKSVEKLSQTDHVLVPSQAADGTQLIQ 177
DB 198 AAIKQDGIITQLTANLQAR-REKDETM-----REFLELTSQSKLQIQFQQLQASETLR 252
QY 178 VASFAFTP-----DKTATAEYTSRAGENGESQLDVGKEIINEGEVFN---SYLLKKVT 230
DB 253 NSTHSSTAADLLQAKQOILTHQOOLEEQDHLE-DYQKKK-----EDFTMQISFLQEKIK 306
QY 231 IPTGYKHIGQDAFVDNKNIAEVLNPSLETISDYAFAPHLAKQIDLPDNL----KAIGEL 286
DB 307 V-----YEMEQDKVENSKEEIOEKEITIEELNTKIIIEEKKTLKDKLTTADKLLGEL 362
QY 287 APFDNQITG-----KLSLP-----RQLMLRAERAFKSNHIKTIEFRGNS 325
DB 363 ---QEIQVQKNQEIKNMKLELTSKQKERSSEIKQLMGTVBELQKRNH-KDSQF---- 414
QY 326 LKVICEASQDNDLSQLMLPD---GLEKIESEFTGNPGDDHYNRVVLWTK----- 374
DB 415 -----ETDIVORMEQETQRKLEQRAEL-----DEMYGOQIVQMKQELIRQHMA 458
QY 375 -----SGKNPSGLAT-ENTYVNPDK-----SLWQSPSIDYTK 406
DB 459 QMEEMKTRHKGEMENALRSYNTVNEDOIKNMVAINELNKLQNTNSQKEKKEELGL 518
QY 407 WLEEDFTYQKN-----SVTGFNSKGLQKVKR---NKNLEIPKQHNGVITTE----- 452
DB 519 ILBEKCALQOEDLVEELSFSREQIQRARTTAEQESKLNZAKHSLSVTEDLKAEIVSA 578
QY 453 NAFR-----NVDFQNTLRYDLEVKLPSTIRKIGAFQSNILKS---FEASDDLE 502
DB 579 SSSRKELELKHAEVNTYKIKLEMLEKERNVLDNRMAESQAELELRQLLFSHEELS 638
QY 503 EIKGAFMNRITETLELKDVLVTIGDAAFHINHIYAI VLPESVQEIGRSAFRQNG----A 558
DB 639 KLKEDLEIHRINIEKLNKL-----GHHYQQIDGLQNEMSQKLTQFQKDNLTQ 692
QY 559 NNLIIFGSKVKTIGEMAFISNRLHLDLSEQKQLTPIPVQAFSDNALKEVLLPASLKTIR 618
DB 693 NQLILEISKLDL-QOOLVNSKSEMTL-----QINE-----LQKEIEILR 732
QY 619 EAFKGNHLKQ-LEVASALSHAFNALDNDGDQFNDKVVKTHNSVALADGEHFIVD 677
DB 733 QSEKEKGTLEQVEQLQELTELLEKQMEKENDLQ-----EKPAQLEAENSILK 781
QY 678 PKLSSTIVDLKILKLEGLDYSLRQTQTOTQFRDMTTAGKALLSKSNLROGEKQKFLQ 737
DB 782 DEKKT-----LEDMLKI-----HTPVQSEERLIFLDSIKS-----KSKDSWSEKEIEILIE 827
QY 738 EAQFFPLGR-VLDLKAIAKAEKALVTTKATKNGOL-----LERSINKA 778
DB 828 ENEDLAQOCIQLNREEIEKQRNTF--SFAEKNFEVNYQELQEEVACILKYKDDLEDSEKKNQ 885
QY 779 VLAYNNSAIIKANVGRKELELDL 801
DB 886 ELEY-----KSKLKALNEELHL 902
```

## RESULT 86

US-10-171-311-6

; Sequence 6, Application US/10171311  
; Publication NO. US20030087270A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Chen, Yan  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Monahan, John  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Glatt, Karen  
; APPLICANT: Gannavarapu, Manjula  
; APPLICANT: Hoerish, Sebastian  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
; FILE REFERENCE: OF CERVICAL CANCER  
; FILE REFERENCE: MRI-035

; CURRENT APPLICATION NUMBER: US/10/171,311  
; CURRENT FILING DATE: 2002-06-12  
; PRIOR APPLICATION NUMBER: US 60/298,159  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,155  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/335,936  
; PRIOR FILING DATE: 2001-11-14  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 3925  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-171-311-6

Query Match 3.6%; Score 185; DB 14; Length 3925;  
Best Local Similarity 19.5%; Pred. No. 0.11;  
Matches 180; Conservative 164; Mismatches 349; Indels 230; Gaps 44;

```
QY 7 TVALTLTVSVTHNQEVSLVKEPIKQT--QASSISGADYAESGSKLKINETS GP 64
DB 82 TMRTLHSGEITSHQG-FSVELESEISTTADDCSEVNGCSFVMRTGKPTNLLRREEFG 140
QY 65 VDDTVTDL-FSDKRTTPEKIDNLAQPREQELKAVTENSEKQITSGSO-LEQSKESL 122
DB 141 VDDSYSEQAQDSPTHELMWSELA--GKQHEIEELNRELE-EMRVYTGEGLOQLQEF 197
QY 123 SLNK-----TVPSTSNWEICDFTKGNLTVGLSKSVEKLSQTDHVLVPSQAADGTQLIQ 177
DB 198 AAIKQDGIITQLTANLQAR-REKDETM-----REFLELTSQSKLQIQFQQLQASETLR 252
QY 178 VASFAFTP-----DKTATAEYTSRAGENGESQLDVGKEIINEGEVFN---SYLLKKVT 230
DB 253 NSTHSSTAADLLQAKQOILTHQOOLEEQDHLE-DYQKKK-----EDFTMQISFLQEKIK 306
QY 231 IPTGYKHIGQDAFVDNKNIAEVLNPSLETISDYAFAPHLAKQIDLPDNL----KAIGEL 286
DB 307 V-----YEMEQDKVENSKEEIOEKEITIEELNTKIIIEEKKTLKDKLTTADKLLGEL 362
QY 287 APFDNQITG-----KLSLP-----RQLMLRAERAFKSNHIKTIEFRGNS 325
DB 363 ---QEIQVQKNQEIKNMKLELTSKQKERSSEIKQLMGTVBELQKRNH-KDSQF---- 414
QY 326 LKVICEASQDNDLSQLMLPD---GLEKIESEFTGNPGDDHYNRVVLWTK----- 374
DB 415 -----ETDIVORMEQETQRKLEQRAEL-----DEMYGOQIVQMKQELIRQHMA 458
QY 375 -----SGKNPSGLAT-ENTYVNPDK-----SLWQSPSIDYTK 406
DB 459 QMEEMKTRHKGEMENALRSYNTVNEDOIKNMVAINELNKLQNTNSQKEKKEELGL 518
QY 407 WLEEDFTYQKN-----SVTGFNSKGLQKVKR---NKNLEIPKQHNGVITTE----- 452
```

Db 519 ILLEKCALQORQLEDLVEELSPREIQIQRARQTIARQESKLNBAHKSLSLTVEDLKAEIVSA 578  
 QY 453 NAFR-----NVDFOKTLRYDLEEVKLPSTIRKIGAFAPQSNLKS---FEASDDLE 502  
 Db 579 SBRSELELKHAEVNTYKIKLEMLEKKNVLDRAWESQAELERLRTQLLFSHEBELS 638  
 QY 503 ELKEGAFMNNRIETLELKDCLVTIGDAPHINHIYAIVLPESVQIGRSAPFONG-----A 558  
 Db 639 KUKDELEIHRINIEKLKDNL-----GIHYKQOQDGLQNEMSQKIETMQFEKDLITKQ 692  
 QY 559 NNLIEMGSKVKTGEMAFLSNRLEHLDLSEQQLTEIPVQAFSDNALKEVLLPASLKTI 618  
 Db 693 NQLILEISKLDL-QOSLVNSKEEWTL-----QINE-----LOKEIEILR 732  
 QY 619 EFAFKKHLKQ-LEVASALSHIAFNALDNDGDEQFDNKKVVVTHHNSYALADGSHFI 677  
 Db 733 QBEKEKGTLEQVQBLQKTELLEKQMKENDLQ-----EKFAQLAEANSILK 781  
 QY 678 PKLSSTIVDLKILKLEGLDYSLRTQTOTQPRDMTTAGKALLSKSNLRQEKQKFLQ 737  
 Db 782 DEKKT-----LEDMLKI-----HTPVQBEERLIFLDSKS-----KSKDSVWEKEIEILIE 827  
 QY 738 EAQFFLGR-VLDJKAIAKAEKALVTCKATKNGQL-----LEERSINKA 778  
 Db 828 ENEDLKQOQIQINIEIEKQRTF--SFAEKNFVNYQELQEBYACLLKVKDDLEDSKNKQ 885  
 QY 779 VLAYNSAIKKNVRLKELDL 801  
 Db 886 ELEY-----XSKLKALNEHL 902

## RESULT 87

US-09-815-242-12610  
 ; Sequence 12610, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Chislen, Karl L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Cart, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; TITLE OF INVENTION: Prokaryotes  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR FILING DATE: 2001-03-21  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12610  
 ; LENGTH: 5795  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus aureus  
 US-09-815-242-12610

Query Match 3.6%; Score 185; DB 9; Length 5795;  
 Best Local Similarity 20.9%; Pred. No. 0.18;

Matches 226; Conservative 142; Mismatches 418; Indels 296; Gaps 50;  
 QY 13 TTVSVVTHNQVFLSVK-----EPILKQTK-----ASSISGADYAESGK 53  
 Db 3817 TTVAGV--NQETAKATELNNNAHSLQNGINDETQTKQKYLDAGPSKKSAYDOAVNAK 3874  
 QY 54 SKLK-----INETSPPVDVTDLFSDDKTTPEKIKDNLAKGPRQELKAVT-----EWTE 104  
 Db 3875 AILTKASQNVYDKAAVEQALQNVNSTKTALNGDAKLAKBAKAAKQTLGLTHINNAORTA 3934  
 QY 105 SEKOITSSOLE-----QSKESLSLNTKVPSTSNWEICD----- 138  
 Db 3935 LDNEILOATNVEGNTVTKAKAQQLDGMQQLETSIRDKDTTLQSONYQDADAKRTAYSQ 3994  
 QY 139 -----FITK-----GNTLVGLSKSGVEKLSQTDHLVLPSSQAADGTQLT-----QVASPAFT 184  
 Db 3995 AVNAAATILNTAGNT-----PKADVERAMQA--VTQANTALNGIQNLERAKQAANTAIT 4048  
 QY 185 -----PDKTAIABYTSRAGENGEBISOLDVDGKEIINEGEVFNISYLLKVTPTPGYKHI 238  
 Db 4049 NASDLNTKQKEALKAQVTSAGRVSAANGVEHTATELNTAMTALKHAIADKAETTKASGNVY 4108  
 QY 239 QODA-----FVONKNIAEVNL-----PESLETISDYAFHL-----ALKQIDLPDLKAI 286  
 Db 4109 NADANKROYDEKVTAAENIVSGTPTPLTPTSDVTNAAQTQVNTAKTQUNGHNHLEVAQON 4168  
 QY 287 AFFDNOITGKLSL-PRQLMRLAERAFKSN-----HIKTIEFRGNSL-----KVIGBAS 333  
 Db 4169 A--NTAIDGLTSLNGPQKAKLEQVGQATTLPNQVTVDNAQTTLNTAMKGLRDSIANEAT 4226  
 QY 334 PQDN-----DLSQLMLPDGLEKIESEAFPTGNPGDDHNNRVVLWTKSGKNPGLATENTYV 389  
 Db 4227 IKADQNYTDSQNKQTD-----YNNAVT-----AAKAIIGQTTTSP-- 4262  
 QY 390 NPDKSLMQESPEIDYTKWLEEDFTYQKNSVTGFSN-----KGLQKVKENKNLE 437  
 Db 4263 -----MNAQEIQAQ--DQVTAQOALNGQENLRTAQTNAKQHLNGLSDLTNAQKDA 4312  
 QY 438 IPKHNGVTTITEIGDNAFRNVDFQNKTLRYKDYLBKVLKLPSTIRKIGAFAPQSNLKSFEA 497  
 Db 4313 AKRQIEGATHVNEVTAQANNADVLNTAM-----TNLKNQIQ-----DQNTIKQGVN 4358  
 QY 498 SDDLBEIKGAFMNNRIETLEKDKLVIGDAAFHINHIYAIVLPESVQIGRSAPFONG 557  
 Db 4359 FTDADAEAKRAYTNAVTAQAEQILNKAQGNPTAKDNVE-----TALQNVQRAKSELNG 4410  
 QY 558 ANNLIEMGSKVKTGEMAFLSNRLEHLDLSEQQLTEIPVQAFSDNALKEVLLPASLKTI 617  
 Db 4411 NQNV-----ANAKTTAK-----NTLNN-----LTSI-----NNAQK-----AALKS- 4441  
 QY 618 REEAFKKNHLKQLEVASALSALSHIAFNALDNDGDEQFDNKKVVVTHHNSYALADGSHFI 677  
 Db 4442 -----QIEGATTVAGV-----NOVSAKATELNTAMSN----- 4468  
 QY 678 PKLSSTIVDLKILKLEGLDYSLRTQTOTQPRDMTTAGKALLSK-----SNLRQEKQK 734  
 Db 4469 ---LQSGIND--EAATKAAQ--KYTDADREKQTAYNDAVTAARTLLDKTAGSDNKAAYEQ 4522  
 QY 735 FLQBAQFPFLGRVDLDKATAKAEKALVTCKATKNGQLLERSINKAVLAYNNSAIKKNV 794  
 Db 4523 ALQ-----RVNTAKTALNGDARLINEAKNTAKQOL-----ATMSHLTNA-QAN-- 4564  
 QY 795 LEKELDLTLGLVBGKGPLAQA--TMVQGVYLLKTPPLPEYVYIGLVNVPYDKSGKLIYALDM 853  
 Db 4565 LTSQIERGTTVAGVQGIQANAGTLDQAMNQLRQSIASKDATKSSSEYQDANADLQNAVNH 4624  
 QY 854 SDTIGEGQKDAYGNPILNVEDNE-----GYHALAVATLADYEGSLDKTILN 900  
 Db 4625 AVSDAEGIIISATNPNPNPDITINQKASQVNSAKSALNGDEKLAQAQKQ-----AKTDI- 4677  
 QY 901 SKLSQLTSTIROV-PTAAVHRAGIFQAIQNAAAAEQLLPKPGT--H--SEKSSSSSESANS 955  
 Db 4678 CHLSDLNNAQOTSATAEVDNAPNLAAVTSKKNKATSLNTAMGNLKHAAEKDNTKRSVNY 4737

```
QY 956 KD 957
Db 4738 TD 4739

RESULT 88
US-10-437-963-106122
; Sequence 106122, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 106122
; LENGTH: 1479
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_10596C.1.pap
US-10-437-963-106122

Query Match 3.6%; Score 184.5; DB 16; Length 1479;
Best Local Similarity 19.1%; Pred. No. 0.029;
Matches 213; Conservative 182; Mismatches 398; Indels 321; Gaps 50;

QY 27 LVKEPLTKQTASSISGADYAESGKS-----KLKINETSQGPVDDT-----VTDL 72
Db 467 LLSERV--QEEAASSAEKLEAEATNSVEAYKEKINELQASLDSTTSKNQLLEQVVDL 524

QY 73 FSKKRTTPKIKDNIAKPREDELKAVTENTSEKQI-----TSGSQLEQSK 119
Db 525 -SDKFTHEQEAHVHE--RSLESLTTSKDAEVAHTRQDLENELNANERKPEVE 581

QY 120 ESLSLNKTVPTSNWEICDFITKGNITVLGSKG-----VEKLSQ-----TDHLVL 165
Db 582 ADLEQYRSKVSQLSDELAAYQTKAASLEAVMESASEKEKELVESLGQITEKKLELLVL 641

QY 166 PQQAADGTQLIQVAS-----FAFTPKKTAIAEYTSRAGE-----NGEISOLDVDGKEIIN 216
Db 642 EYBEKTEEYVKESQLEELQESKVLALBESLVKGEESHEGTIADNL-----QLSN 698

QY 217 EGEVFNYSLLKKVTIPTGYKHIGQDAFVDNKNIAEVLN--PESLETISDYAFALHAKQI- 274
Db 699 KNDY-----MLESQLSQAGDDHSKTRSLLSQAHLKHELE-----LNLKSLN 742

QY 275 DLPDNLKATGELAFPDNQTGLSLPQLMRLAERAFKSNHKTTFPRGNSLKVTGEASF 334
Db 743 DLHVESKTAESAL---ORIALETQVQELSAAEQSLKS--HLTEFESK-----LASAEK 792

QY 335 QNDLSQLMPLGELKIES-----EAFPGDGDHNNRVVL---WTKSGKNPSGLATEN 386
Db 793 KSMDEQ--ELKATECNSRLRVDELSEL--EYKKEKTSLEASLEAKQEAELSEKL 850

QY 387 TVYNPKSLQWSPSEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKNLEIPKONGVT 446
Db 851 DQWNEEKFEELSKKATIKHLEA-----NQVQAL--QGELESARHKLVEESDLQAL 903

QY 447 ITEIG-----DNAFRNVDFQNTLRYDLEEVKLPSTIRKIGAFAPQSNLKSFPASDDL 501
Db 904 IRETSLDKLGAEEQLEHKGALHATSKKIDLEALYQSL-----LEDTEMKLOQAGENL 959

QY 502 EEIKEGAFMNNRIETLELKDKLVITGD--AAFHINHIYAIVLPSVQVIGRSAPRONGAN 559
Db 960 TQ-----KETECQELSEKLAAEQAAASYQAKATAAAAEVESV--KVELEAFE----- 1005

QY 560 NLIFMGSKVKTIL--GEMAFLSNRLHLDLSEKQLTETIPVOAFSDNALKEVLLPASIKTIR 618
Db 1006 -----TEISTLETTIEELTKASNAESRAEQALVESAMMSETNQALKED--LDKALMLUR 1058

QY 619 E--EAFKKNHLKQLEVASALS-----FDMKNVVKVTHNSYALADGHEFIVDDP-----KLSSTIVDLK 690
Db 1059 ELQEQFDSHAEKEEVFTKLSAHEKTIHLETVHSGRGLHATAESKNAELEAQHAELE 1118

QY 643 ALDDNDGDEQ-----FDMKNVVKVTHNSYALADGHEFIVDDP-----KLSSTIVDLK 690
Db 1119 TIGKKDSVKDLNRLAALAESEIESLTHVNEAMKQEIINAKLVKVDLEQELKLSISSISSEKEE 1178

QY 691 ILKIEGLDYSLTROTTOFRDMWTAGKALLSKSNLRQGE-----731
Db 1179 VAEKV-----VVHEKTTIEHLREHSHRGLQSAAESRSAAETENELREVLETVAQKEARV 1232

QY 732 ---KQKFLQ---EAOFFLG-----RVDLDKAIA-----KAEKALVTK 762
Db 1233 TDLKEKLVSTENKLVINEALKGELDTKVAMPDELOEQEPPSSHAEKEEAELKLVH 1292

QY 763 K-----ATKNGQLLERSINKAVLAYNNSAIKANKVRLKLEKLDLL-- 802
Db 1293 RTISHLTEVHTRSLELSAESAESKNEEIESKLHEAL--EMAQKEAEVKDL SKLDLEI 1349

QY 803 -TGLVEGKPLAQATMVQGVLLKTPPLPEYIGLVNLYFDKSGKLIYALDMSDITIGEQ 861
Db 1350 ELGYTTEQA--TEAAAABETHKIK-----FDEAVHKIKSLEQLAVTENK 1392

QY 862 KDAYGNPILNVDENEGYHALAVATLADYEGLDIKTILNLSKLSOLTSIRQVPTAAVHRAG 921
Db 1393 VE-----LPFTEKEN-----LVIANSKLNEELHLQ-----N 1419

QY 922 IFQAIQNAARAEQALLPKPGTHSEKSSSESANS 955
Db 1420 KNLQVALAAV-----AEKEGSGSEIHS 1444

RESULT 89
US-10-369-493-13477
; Sequence 13477, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13477
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Thermoplasma volcanium
US-10-369-493-13477

Query Match 3.6%; Score 184; DB 15; Length 891;
Best Local Similarity 19.5%; Pred. No. 0.015;
Matches 198; Conservative 159; Mismatches 335; Indels 322; Gaps 52;

QY 51 SKSKLKNETSQPVDDTVTDLFSDKRTTTPKIKDNIAKPREDEL-----96
Db 34 AGKSSI-----VDAIRFALFSDKRT--RRTEDMIKKGERYMEVELYFRSEGHYSYIR 83
```



```

QY 97 -----KAVTENTESEKQ---ITSGSQLEQS---KESISLNTKVPSTSNWEICDPITKG 143
Db 84 RTIERGKSISTDASIERDGSIIITGASDVSNVKNVNLINIKOVFLTS-----IFVRQG 138
QY 144 NTLVGLSKSGVEKLSQTDHLVPSQAADGTQIQVASFAPTPDKKTAIAEYTSRAGE--NG 202
Db 139 EMDALVSKDPAERKKILDEILNIDRLAAGYLLK---EVIDDLTANVSDYDYLKNEIQLS 194
QY 203 BISQJVDGKEI-----INEGVFNYSYLLKKVTIPTGYKHICQDAPVNNKNAEYVL 254
Db 195 KINEIDNNKQTEELSKRLIEPEI--KALEEINIKENKK-----DHLNEELHRL 244
QY 255 PESLETSIDYAF-----AHLAKQIDLP---DNLKAIGELAFF--DNOITGKLSLP 300
Db 245 NAQLETIKKYEMLAESRKSASIEWVVKGPSIEBELKRLNNAVVKRNEIIEYINLK 304
QY 301 RQLMLAERAFKSNHIKTIERFNGSLKIVGEA-----SPQNDLSQLMLPGCLEKIE 352
Db 305 KDLGSLSE-----IEGLKSDLKYDEAHRKLEDLQSPRSEFFLEKKRKEKEDLKL 355
QY 353 SFAFTGNPGDDHNNRVVLTWYKSGKNSGLATENTYVNPDKSLMOESPED--YTKWLEED 411
Db 356 SSL-----KEDEONYQSAV-----RNIEIKKWIENEKEIERMSAFISEI 396
QY 412 FTYQK-----NSVTGFSNKGIGQVK-----RNKNLEIPKQ-----441
Db 397 LKIQITPEILNRRRAEINSSLMQIEGKIASINASIDAMRSHKMEVEENAAWLSRGVCP 456
QY 442 -----HNGVTITE-----IGDNAFRVNDPQNTK---LRKYD-----LBEVKLPSTIR 480
Db 457 VCGTHLGTEKSEDLVKHYGEASRLSEEDINKTENEIKKLBDEERKHQKLLDRINGKOVER 516
QY 481 KIGAFAPQSN---NLKSPASDDELEIEGAP-----MNRITELBKOKLVITIGDAA 530
Db 517 LIATSNLLSSKRAELKKEF--MDDEARLKEAHLKAAEAI SQYNSIDLGLEAK---NEEW 570
QY 531 PHINHIYAIVLPESVQIEGRSAFRONGANNLI FMGSKVKTLCGEMAFSLNRLHLDLSQK 590
Db 571 LKANAVISDIENI-----RSRFE-----KNQNLNIIKRWNEIE-----607
QY 591 QLTTRIP--VQAFSDNALKEVLLPASLKTIREBAFKKNHL-----KOLEVASALSHTAFNAL 644
Db 608 ---VNPIDVESYNENSLKRI--DBELNSLRN---KNELYAKKAAWDEIQKTIEHF-----655
QY 645 DNDGDGDFDNKVVVTKHNSVALADGHEFIVDPD--KLSSTVDLEKILKILIEGLDYST 702
Db 656 -----KEBISKK---KGIEDSOAEVNAQLQINDDLKQLSSR---LDKI--NVDQYEWKS 702
QY 703 LRQTTQTQFRDMMTAGKALL---SKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKAL 759
Db 703 LH-----KVLLOQNEKLNIAVADIRKRLEKKEITII-----KAIADLKR--740
QY 760 VTKKATKNG--OLLERSINKAVLA YNNSAIKANKVRLKLEKLD-----LLTGL 805
Db 741 VREAFSKDGVPAI IRKSASEPI---TNQT--RQYIQR FELDIDDVDVDQDFNITVFRGCI 795
QY 806 VEGKOPLAQATWQVQVYLLKPLPEYIYGLNVYFDKSGKLIYALDMSDITIGEGQKDAY 865
Db 796 AEGIDLSGGERMAVAFALR-----VAIAQFLNKDVLSL-----VM 831
QY 866 GNPILNVDEDEGYHALAVATLADYEGLDIKTLNSKLSQLTSIRQVPTAAVHR 919
Db 832 DEPTAFLEDDRSS-----DLANIIEYLSKDSGGIPQVIMISHRR 870

```

## RESULT 90

```

US-10-282-122A-71690
; Sequence 71690, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangueu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl

```

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71690
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-71690

```

```

Query Match 3.6%; Score 184; DB 15; Length 1665;
Best Local Similarity 19.4%; Pred. No. 0.036;
Matches 210; Conservative 186; Mismatches 430; Indels 258; Gaps 50;

QY 16 SVVTHNOEVSLVKEPILKQTAQSSISGADYAEBSGSKLKNETSGVDDTV--TDLF 73
Db 365 NVINNTTEATEERKQVALTKLN-----KYNEKKEIDNLSSSDIA 406
QY 74 SDKRTTPEKIDNLAQPREQELK-----AVTEN-----TSEKQITSGSQLEQS 118
Db 407 PLKESSIAQINSINVRATKAAAKQAIEAALTDRKVFIDSHYDATQEKDVAMAKVTSEA 466
QY 119 KESISLNTKVPSTSNWEI-----CDFITKGNLTVLGSSGVEKLSQTDHL 163
Db 467 NKAKAL--IDQSTSNNDVDQAQTNGINIINSIDADVTKKAN-----AKKAIEQAEEAKKA 519
QY 164 VLPQQAADGTQIQVASFAPTPDK---KTAIAEYTSRAG---ENGEISOLDVDGKEII 215
Db 520 LI--NQNSDATQEEKDAAIQRTVDEVNRADRLIDQSTNNDGVDEVOAHSSISNINIQEIV 578
QY 216 NEGVRFNYSYLLKKVTIPTGYKHICQDAPVNNKNAEYVLPSLETSIDYAFAPHL--ALUK 273
Db 579 KKSDA-----KQAI DTAV--LNQKSLVNNN-----EATQEKDVALAKIDEAAKQ 622
QY 274 ----IDLDPDNLKAIGELAFPDNQITGKLS--LPQLMLAERAFKSNHIKTIERFNGSLK 327
Db 623 AKAAIDAATTNNAVDEAT---NNNTTIISGLPDTVVKAAAR-----KAIDAAATK 672
QY 328 VIGBASFDQNDLSQLMLPDGLEKIE-----SEAFNGNPGDDHNNRVVLTWKS GK 377
Db 673 ---EAINNTSDATQEEKDAALAKYAAVTAAKQAITQATTNDNDVQEQN-----718

```







Db 663 DAEQALSATNGNHLSDKSELKQESD-----MEDIDGANLCADPGVVVPALHGETTSSDLAD 717  
Qy 671 GEHFIV-----DPDKLSSTIVDELEKILKILIEGLDYSTLRQTTQTPR 712  
Db 718 NDGAEVENSTPACDGLASSGAPTGNDSKNSAAAQAQVEEDVPSQDDAQQVEEDVPSQDD 777  
Qy 713 DMTTAGKA--LLSKNLNRQGEKQKFLQEAQFPLGRVOLDKATAKAELAVTKATKNGOL 770  
Db 778 DNPADGAPGEICSNANAFSTSSCAVEYV-----QDIASVTVDIHDKNDND--- 827  
Qy 771 LERSINKAVLAYNNSAIKANVKRLKELDLTLGLVEGKGPLAQATMVQGVYLLKTPPL 830  
Db 828 -DENINTDITG-NHSEPK-----LETNVD-----NEDRGDI-QVKKPVPVLMKVP--- 870  
Qy 831 PEYIYGLNVYFDKSKGLIYALDMSDTIGEGQKDAYGNPILNVDEBNEGYHALAV----- 884  
Db 871 -----RPMSESHWEKIQDAQ-----ICDELTKRDAINVLRQKK 906  
Qy 885 ATLADY-----EGLDIKTLNLSKLSQLTSSIRQVPTAAVHRAGIQAIQNAARAAEAQ 935  
Db 907 ALCDYREKLEAAROEERGAFTAHGDKNDLNSVQSM-IGRMNTRANSIQEIDDMTAMKEK 965  
Qy 936 LUPKPGTHSEKSSSESANSK-DRGLQSNPKTNRGRHSAILPRTGSK 981  
Db 966 II-----AHESI5LKEKELLQDIKELKAKQKQLSSNWGSK 1001

## RESULT 95

US-09-815-242-11391  
; Sequence 11391, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kazi L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11391  
; LENGTH: 944  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
US-09-815-242-11391

Query Match 3.6%; Score 182.5; DB 9; Length 944;  
Best Local Similarity 19.2%; Pred. No. 0.02;  
Matches 201; Conservative 148; Mismatches 327; Indels 371; Gaps 49;

Qy 6 KTVALTITVSVVTHNQ--EVFSLVKEPILKQTQASSISGADYAESGSKLKLNETSG 63

Db 29 KDIGLELKTNSKMTPEQAGKLYKIVDGIKEQIQANQP---AKNPQDNKDDLATAVASK 85  
Qy 64 PVDVTVDLPSDKKT---TPEKJKDN-----LAGPREQELKAVTEN-----TESBKQ 108  
Db 86 SLNKKVSKTPKKEETKSPQPKTKKEKKGAAPTIAKKKGGEIIVNTFENQTPPTENTPK 145  
Qy 109 ITSGSOLEQSKESLSLNKTVSTSTNWEICDPIITKGTNTLVGLSKSGVEKLSQTDHLVLPQ 168  
Db 146 VVSHSQIEKAKQKIQ-----EIQKSRREALNKLTSN----- 176  
Qy 169 AADGTQLIQVASFAPTPDKKTAIAEYTSRAGE-NGEISQLDVDGKEIINEGEVENSYLK 227  
Db 177 -----ANNASNANNAKKEISEVKKQEQEIKRH-----E 204  
Qy 228 KVTIPTGYKHITGQDAFVONKNIAEVNLPSELETISDYAFAHLAKQIDLDPNLKAIGELA 287  
Db 205 NIKERTGPRVIRKNDENEVE--SENSVTESKKPTQSAAAIPEDIKK-----EWQ 251  
Qy 288 PFDNQITGKLSLPLQMLRAERAPKSNHIKTI-----EPRGNSLKVTIGEASFQDNLSQ 341  
Db 252 EKDQEAQKAKKPSK--PKATPTAKNNKSHKIDFSDARDPKGNDI-----YDDEYDE 301  
Qy 342 LML-----PDGLEKIESEAFNGDDHNNRVLTWKSGKNPSGLATENTYVNPDKSLW 396  
Db 302 ILLFDLHEQDNFKNEEBE---KSIQRQINDRVV---QRKNP----- 337  
Qy 397 QESPEIDYTKWLEED-FTYQKNSVTGFSNKGLOKQVKNKNLEIPKQNGVTTITIGDNAP 455  
Db 338 -----WMNESGIKQSKKKRAFRNDNSQKVIQS-TTAIPEE---VRVYFAQKAN 383  
Qy 456 RNV-----DFQNTLRKYDLEEVKLPSTIRKIGAFQFQNNLSKSPAS 498  
Db 384 LNLADVITKLPNLGLMVTNDFDKDSIEILAEFHLSEISVQ-----NTLEEPEVE 434  
Qy 499 DDLBEIKE-----GAPMNNRIETLE-LKDKLVTIGDAAPHINHIYAVLPES--- 544  
Db 435 EVLEGVKKERPPVVTIMGHVDHGKTSLLDKIRDKRVAHTEAGGITQHIGATMVEKNDKW 494  
Qy 545 --VOEIGRSAP--RONGAN---NLIPM---GSKVKTIGEM-----AFLSNRL 581  
Db 495 SFIDTPGHEAFSOMNRGAQVTDIAVIVIAADGCVKQQTIEALEHAKAANVPVIFAMNM 554  
Qy 582 EHLDSQOKLTE-----IPVQAFS-----DNALKEVLLPASLKTREE 620  
Db 555 DKPNVNPDKLKAECALGYNPVDMGGEHEFTIPVSAKTGDGIDNLETLIQAGIM----- 609  
Qy 621 AFKNHLKOLBVASALSASHIAFNALDDNDGDSQFNDKVVVKTTHNSYALADGEHFTVD--- 677  
Db 610 -----ELKAIIEGSAARAVVLEGSVEKGRG-----AVATVIVOSGTLVSUGDSFAETAF 657  
Qy 678 -----PDKLSSTIVDLE-KILKLEIGLDYSTLRQTTQTPFRDMTTAGKALLSKSN----- 726  
Db 658 GKVRTMTDDQCKSIQNLKPSMVALITGLS-----EVPAGSVLIGVENDSIA 704  
Qy 727 -LROGEKQKFLQEAQFPLGRVOLDKATAKAKA-----LVTKATKN-----G 768  
Db 705 RLQAKGKATYLRQ-----KALSKSTKVFDSELMVANKELKNIPVVKADTQ 753  
Qy 769 QLLERSINKAVLAYNNSAIKANVKR-----LEKELDL-----TGLVEG 808  
Db 754 SL--EAIKNSLLELNNEEVALQVHSGVGGTENDLSLVSSSEHAVILGFNIRTPGNVK 811  
Qy 809 KGPLAQATMVQGVYLLKTPPLPEYIYGLNVYFDKSKGLIYALDMSDTIGEGQKDAYG-- 866  
Db 812 KA-----KEYNVSIKTY-----TVIYAL-----IEBMSRLILGLM 841  
Qy 867 NPILNVD-----EDNEGYHALAVATLA 888  
Db 842 SPIIEEHTGOAEVRETFNIPKVGTTIA 868

RESULT 96





```
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10903
/ LENGTH: 1192
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
/ US-09-815-242-10903

Query Match      3.6%; Score 182; DB 9; Length 1192;
Best Local Similarity 19.0%; Pred. No. 0.03;
Matches 218; Conservative 177; Mismatches 431; Indels 324; Gaps 51;

QY 46 DYAES-----CKSKLNKINETSQVDDVTDTLPDSK-----RT 78
Dy 94 DYSEISVTRRLKRTGESDFFINKQACRLKD-IQDLFMSGSLGKESFSISQKVEAIFNS 152
QY 79 TPEK-----IKONLAKGPREQELKAVTENT-----ESEKQITS-GSQLEQS 118
Dy 153 KEDRRGIPPEAAGLVKYQKQKKAQKLFEPEDNLRSQVDIIYLEQLVPLAQADA 212
QY 119 KESLSL-----NKTVPSTN-----WEICDTITKNTLVGLSKSGVEKLSQTDHVLV 165
Dy 213 KKLALKBELTEIDVNLTVTEIQEAKAIWE-----TKQELTAIE-----EKLAKASKQVH 263
QY 166 PQAADGTGLIQVAFPTDKKTAIEYTSRAGENGESQLDVGKELINEGEVFNPSYL 225
Dy 264 DLEG-----KLVRLS-----KXRLDE-----QIETEQQLQVTEALKQAE 301
QY 226 LKKVTIPTGYGHIGQDAFVNDKNIAE-----VNLPELETISDYAFAPHLAKQIDLPNL 280
Dy 302 GQKVLERSKHTSQTASBYETLAETAEKIVRYREELQTLTAETAKTAQKQ-----TL 356
QY 281 KAIGELAFDQNTOKLSLPLQMLR-----LAERAFKSNHIKTIEFR-----GNS 325
Dy 357 KEALALATKDVKEYSKSS--KELMBELRSQYEVVMQEQANTANDLKYLERQYQBSTAKNQ 414
QY 326 LAVIGEASFQDNDLSQMLPDGLEKIESFAFTG--NPGDDHYNNRVVLWTKSGKNPSGLA 383
Dy 415 QSLAKHEALDEQMVLEALAMKETLEKEQKVAQLOEQLEEEYFALKATLEAK-----465
QY 384 TENTYVNPDKSLWQESPEIDYTKWLEEDPTYQKNSVTGFSNKGLOKVKRNKN-----435
Dy 466 -RERLAQRQNDYQAMNQVQAKARQKSLQELQENYAGF-YQGVKAVLHKKQLTGIVCA 523
QY 436 -----LEIPKHQNGVTITEIGDNFRNVDPNQNTLKYDLLEVK-----LP-STIRK 481
Dy 524 VAELEVPKREYTLATETALG-GAAQHI VVENEKDRAGITFLKQOHSGRATEPLPTTIKP 582
QY 482 IGAFAFQSNKLSFEA-----SDDLSEIKE--GAFMNNRIETLEKDKLVITIGDAAPHNH 535
Dy 583 RVSAMVQNRLAGAGFVGIASELVRYPEQVQVTIQLNLGVTILAAIDLTSANQLAKLVNY 642
QY 536 IYAIIV-LPESVQIEGRSAFRQNGANNLIPMGSKVKTTLGEMAFLSNRLHDLIDSEKQLTE 594
Dy 643 QYRVVSLGEGVNVPGGS--MTGANKRGNGQSLFSAQELQITITQMTQLE--TQLRSVE 698
QY 595 IPVQAFSDNALKEVLLPASLKTIRBEAFKKNHLKQLEVASALSHIAFNALDNDGDGEQFD 654
Dy 699 QEVQALS-----QEVKTATERAEMLSRAGQNRKLQKEI-----D 733
QY 655 NKVVVYKTHNSVALADGEHFIVDPDKLSSTIVDLEKILKLEIG-----LDYSTLRQ 705

/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10903
/ LENGTH: 1192
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
/ US-09-815-242-10903

Query Match      3.6%; Score 182; DB 9; Length 1192;
Best Local Similarity 19.0%; Pred. No. 0.03;
Matches 218; Conservative 177; Mismatches 431; Indels 324; Gaps 51;

QY 46 DYAES-----CKSKLNKINETSQVDDVTDTLPDSK-----RT 78
Dy 94 DYSEISVTRRLKRTGESDFFINKQACRLKD-IQDLFMSGSLGKESFSISQKVEAIFNS 152
QY 79 TPEK-----IKONLAKGPREQELKAVTENT-----ESEKQITS-GSQLEQS 118
Dy 153 KEDRRGIPPEAAGLVKYQKQKKAQKLFEPEDNLRSQVDIIYLEQLVPLAQADA 212
QY 119 KESLSL-----NKTVPSTN-----WEICDTITKNTLVGLSKSGVEKLSQTDHVLV 165
Dy 213 KKLALKBELTEIDVNLTVTEIQEAKAIWE-----TKQELTAIE-----EKLAKASKQVH 263
QY 166 PQAADGTGLIQVAFPTDKKTAIEYTSRAGENGESQLDVGKELINEGEVFNPSYL 225
Dy 264 DLEG-----KLVRLS-----KXRLDE-----QIETEQQLQVTEALKQAE 301
QY 226 LKKVTIPTGYGHIGQDAFVNDKNIAE-----VNLPELETISDYAFAPHLAKQIDLPNL 280
Dy 302 GQKVLERSKHTSQTASBYETLAETAEKIVRYREELQTLTAETAKTAQKQ-----TL 356
QY 281 KAIGELAFDQNTOKLSLPLQMLR-----LAERAFKSNHIKTIEFR-----GNS 325
Dy 357 KEALALATKDVKEYSKSS--KELMBELRSQYEVVMQEQANTANDLKYLERQYQBSTAKNQ 414
QY 326 LAVIGEASFQDNDLSQMLPDGLEKIESFAFTG--NPGDDHYNNRVVLWTKSGKNPSGLA 383
Dy 415 QSLAKHEALDEQMVLEALAMKETLEKEQKVAQLOEQLEEEYFALKATLEAK-----465
QY 384 TENTYVNPDKSLWQESPEIDYTKWLEEDPTYQKNSVTGFSNKGLOKVKRNKN-----435
Dy 466 -RERLAQRQNDYQAMNQVQAKARQKSLQELQENYAGF-YQGVKAVLHKKQLTGIVCA 523
QY 436 -----LEIPKHQNGVTITEIGDNFRNVDPNQNTLKYDLLEVK-----LP-STIRK 481
Dy 524 VAELEVPKREYTLATETALG-GAAQHI VVENEKDRAGITFLKQOHSGRATEPLPTTIKP 582
QY 482 IGAFAFQSNKLSFEA-----SDDLSEIKE--GAFMNNRIETLEKDKLVITIGDAAPHNH 535
Dy 583 RVSAMVQNRLAGAGFVGIASELVRYPEQVQVTIQLNLGVTILAAIDLTSANQLAKLVNY 642
QY 536 IYAIIV-LPESVQIEGRSAFRQNGANNLIPMGSKVKTTLGEMAFLSNRLHDLIDSEKQLTE 594
Dy 643 QYRVVSLGEGVNVPGGS--MTGANKRGNGQSLFSAQELQITITQMTQLE--TQLRSVE 698
QY 595 IPVQAFSDNALKEVLLPASLKTIRBEAFKKNHLKQLEVASALSHIAFNALDNDGDGEQFD 654
Dy 699 QEVQALS-----QEVKTATERAEMLSRAGQNRKLQKEI-----D 733
QY 655 NKVVVYKTHNSVALADGEHFIVDPDKLSSTIVDLEKILKLEIG-----LDYSTLRQ 705

/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10903
/ LENGTH: 1192
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
/ US-09-815-242-10903

Query Match      3.6%; Score 181.5; DB 9; Length 1786;
Best Local Similarity 18.4%; Pred. No. 0.057;
Matches 204; Conservative 205; Mismatches 413; Indels 289; Gaps 47;

QY 22 QEVSLVKEPILKQTOASSSISGADYAESKSKLKNETSGPVDVTDVTLFSDKRTTPE 81
Dy 781 ESVANVEESVAPVVEEIVAPSVESVAPSVESVAENVATNLDNLNLGGIET--E 838
QY 82 KIKONLAKGPREQELKAVT-----ENTESEKQITSGSQESKESLSLNTKVPSTSNWE 135
Dy 839 EIKOSILNEIEVENVVTILENVEETAEBSVTFSSNILEIQENTITNDTIE-----892
QY 136 ICDFTITKNTLVGLSKSGVEKLSQTDHVLVPSQAADGTQLIOVASFAFTPKKTAIABYT 195
Dy 893 -----EKLELHENVL--SAALENTQ-----SBEKKEVIDVI 923

/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10903
/ LENGTH: 1192
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
/ US-09-815-242-10903

Query Match      3.6%; Score 181.5; DB 9; Length 1786;
Best Local Similarity 18.4%; Pred. No. 0.057;
Matches 204; Conservative 205; Mismatches 413; Indels 289; Gaps 47;

QY 22 QEVSLVKEPILKQTOASSSISGADYAESKSKLKNETSGPVDVTDVTLFSDKRTTPE 81
Dy 781 ESVANVEESVAPVVEEIVAPSVESVAPSVESVAENVATNLDNLNLGGIET--E 838
QY 82 KIKONLAKGPREQELKAVT-----ENTESEKQITSGSQESKESLSLNTKVPSTSNWE 135
Dy 839 EIKOSILNEIEVENVVTILENVEETAEBSVTFSSNILEIQENTITNDTIE-----892
QY 136 ICDFTITKNTLVGLSKSGVEKLSQTDHVLVPSQAADGTQLIOVASFAFTPKKTAIABYT 195
Dy 893 -----EKLELHENVL--SAALENTQ-----SBEKKEVIDVI 923
```



```
Db 1454 SLEDT-ITLGERVESLKDVLSSALGMBEEMKTKKQAQ-----RPLKEVLLKEVKEEP 1508
Qy 757 KALVTYKATK-----NGQLLERSINKAVLAYNNSAIKKANVKLEKELDILT 803
Db 1509 KKKITKKVRFDKDKPKDKEIVEEMKDDEIDEDVEEDIEEDKVEDIDEDDEDI 1568
Qy 804 GLVEGKGPLAQTM-----VQGVYLLKTPL--PLPEYIGLVNYPDKSGKLIYALDMSDT 856
Db 1569 G--EDKRDVIDLIVQEKRIEYKAKKKLEKKVEGVSGLKHGVDEVMKYVQKIDREVD 1626
Qy 857 IGEQKDAYGNPILNVDENEGYHA-----LAVATLADY----- 890
Db 1627 KEVSKALESKNDVTNVLQNDQFFSKVNFVKKYKFAAPFISAVAAFASTVGVFTFSL 1686
Qy 891 ----EGLDIKTLNSKLSQ--LTSIRQVPTAAVHRAGIFQAIQNAAAAEQQLPKPGTHSE 945
Db 1687 FSSCVTIASSTVLLSKVDKTNKNERPFYSF-----VDFIFKNLKHYLQOM-----KE 1735
Qy 946 KSSSESAN-----SKDRGLQSNPKTNR 968
Db 1736 KFSKEKNNVIEVTNKAEBKGNVQVNTNTEK 1766

RESULT 100
US-09-801-368-86
; Sequence 86, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 86
; LENGTH: 2026
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; US-09-801-368-86

Query Match 3.6%; Score 181.5; DB 9; Length 2026;
Best Local Similarity 20.1%; Pred No. 0.068;
Matches 223; Conservative 145; Mismatches 365; Indels 375; Gaps 48;

Qy 29 KEPILKOTQA-----SSSI--SGADYAESGSKLKINETSGPV----- 65
Db 616 KTPILNKTSYTKKFTSSSVNNSPDGAQSSG--LLQDEKXDEVECCLEHYKDFSDLD 673
Qy 66 -----DDTVDLFSDKRTTPEKIDNLAGPREBELKAVTENSEKQITSGS 113
Db 674 PKRHYAIRFNTDDFTTLLSCTPATVVEIIPALKR-----KFNITAQG 717
Qy 114 QLEQSKESLSLNKTVPSNNEICDFTKNGTLVGLSKS-----GVEKLSQT-----DHL 163
Db 718 NFQISLKGKLSKILRPTSKPILIE--RKLILLNGYRKSDDPLHMGIEDLSFVFKFLPH 775
```

```
Qy 164 VLPQAAADGTOLIQVAFAPFPDPKKTAAIAEYTSRAGENGESQLDVOCKEINNEGEVNS 223
Db 776 VTPSH-----FTPEQORIM-----RSEFVHVDLRNMDLTTPPIIF-- 811
Qy 224 YLLAKVITPTGYKHIGQDAFVDNKNIAEAVNLPESLETISDYAFAPHAHALAKQIDL----- 276
Db 812 -----YQHTSEIESLOVSNANNIFLP--LEFI--ESSIKLSLRMWNIRASKEPS 857
Qy 277 -----PDNLKAIGELAFPDNQITGKLSLPR-----QLMRLA 307
Db 858 NITKAYKLVSLQRLQNFIRKVPNSIMKLSNLTILNQCNELESIPAGFVELKNLQLLDLS 917
Qy 308 ERAP-----KSNHIKTI-----BFRNSLKVIIEASPOD 336
Db 918 SNKEMHYPEVINYCTNLLQIDLDSYNKIQSLPQSTKYLVKLAKMNLSHKNLFIIDLS--EM 976
Qy 337 NDLSQLMLPDGLEKIESEAFNGPDHNNRVLVWTKSGKNPSGLATENTYVNPDKSL-- 395
Db 977 TDLRTLNL-----RYNRISIKTNASNLQNLFLTNDNISFEDTLP 1017
Qy 396 -----WQESPEIDYTKWLEEDFTYQKNSVTGFNSKG-----LOKVKNK 434
Db 1018 KLRALIEQENP---ITSISFKDF-YPKNMTSLTLNKAQLSSIPGELLTKLSFLEKLELNQ 1073
Qy 435 N--LEIPKQHNGVITTEIGNAFNRVDF-----ONKTLRKYDLEEVKLPSTIRKIGAPA 486
Db 1074 NNLRLPQEISKLTCLVFLSVARNKLEYIPPELSQLKSLRTLDL----- 1117
Qy 487 FQSNLASF--EASDDLE-----EIKEGAPMNNRIETLEKDKLVITGDAAFHNIHAIVL 541
Db 1118 -HSNNIRDFVGMENLELTSLNISSNAPGNSLEN-----SFYHNSYG--- 1160
Qy 542 PESQVEIGRSFQRONGANNLI-----FMGSKVKTLCGEMAFSLNRLEHLDLSEOKQ 591
Db 1161 -----SKLSKSLMFFIAADNQDADAMPFNCVFNKLVNLNLYNNP--SDVSHMKL---BS 1211
Qy 592 LTEIPVOAFSDNAL-----KEVLLPASLKTIREPAFKKHHLKQLEVASALSH----- 638
Db 1212 ITEL---YLSQNKLTLLSGDTVLKWSLSLKTIL-----MLNSNQMLSLPAELSNLSQLSVFD 1263
Qy 639 IAFNALDDNDGEQFD-----NKVVVTHNSYALADGEHFI---VDPDKLSSTIVDLEK 690
Db 1264 VGANQLKYNISNYHYDNWNRNKNKELYNFSGNRFRFEIKFSIHDIADLSDLTVLPLQK 1323
Qy 691 ILKLIIEGLDYSTLRTQTQTQTFDRMTTAGKALLSKSNLRQGEKQKFLQEAQPFGRVDLDK 750
Db 1324 VLGMDVTLNTTKYPDENNVNRLFTTASII---NGMRYGVADTLGQRDYVSSRDVTFER 1379
Qy 751 AIAKAEKALVTKKATKNG-----QLER----- 773
Db 1380 FRGNDECLLCHDSKQONADYGHNISKIVRIDYDKILIRQLERYGDETDNNTALRFS 1439
Qy 774 --SINKAVLAYNNSAIKKANVKLEKELDILT-----LVEGK-----GPLAQTM 817
Db 1440 FLQLNKEINGMLNSVDNGADVNL--SYADLLSGACSTVIYIRGKGLFAANLGDGMALSK 1498
Qy 818 VQGVYLLKTPPLP-----EY-YIGLVNYPDKSGKLIYALDMSDTIG--EGQKDAYGNPIL 870
Db 1499 NNGDYQTLTKQHLPTKKEEYERIRISGYVNGKLDGVVDVSRVGFDDLPHIHASFDI 1558
Qy 871 NVDEBNEGYHALAVATLADYEGLDIKTI 898
Db 1559 SVVTLTKADEMLIIVATHKLMEYMDVTV 1586
```

Search completed: August 28, 2005, 11:02:39  
Job time : 220 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:40:26 ; Search time 45 seconds  
(without alignments)  
1672.139 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 5080

Sequence: 1 MKKHUKTVALTITVSVVTH.....LOYTSVALLSLTAIKKKY 1008

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 110 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*

2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*

3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*

4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*

5: /cgn2\_6/prodata/1/1aa/PCUS\_COMB.pep:\*

6: /cgn2\_6/prodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	215.5	4.2	641	4	US-09-583-110-4263
2	215.5	4.2	641	4	US-09-107-433-2948
3	214.5	4.2	3259	4	US-09-949-016-6507
4	213	4.2	1799	4	US-09-134-000C-5178
5	205	4.0	5024	4	US-09-710-279-2964
6	205	4.0	10182	3	US-09-134-001C-3159
7	204.5	4.0	324	4	US-09-248-796A-18798
8	199.5	3.9	2285	3	US-09-308-375-2
9	199.5	3.9	2285	4	US-09-932-183A-2
10	196.5	3.9	1588	5	PCT-US93-07261-11
11	196.5	3.9	1663	5	PCT-US93-07261-16
12	195	3.8	3896	3	US-09-134-001C-5080
13	192.5	3.8	2482	1	US-08-328-254-6
14	191.5	3.8	3248	1	US-08-353-700-1
15	191.5	3.8	3248	5	PCT-US95-16216-1
16	187.5	3.7	3210	4	US-09-538-092-1154
17	185	3.6	3878	4	US-09-914-259-11
18	183	3.6	1208	4	US-09-134-000C-5756
19	181.5	3.6	1786	3	US-08-973-462-8
20	179.5	3.5	928	4	US-09-134-000C-6590
21	178	3.5	1224	4	US-09-107-433-4347
22	177.5	3.5	1427	4	US-09-538-092-1044
23	176	3.5	1430	3	US-09-008-172-2
24	176	3.5	1430	3	US-09-210-361-6
25	176	3.5	1430	4	US-09-740-274-6
26	175.5	3.5	1078	4	US-09-248-796A-20284
27	175.5	3.5	1216	4	US-09-583-110-3824
					Sequence 4263, Ap
					Sequence 2948, Ap
					Sequence 6507, Ap
					Sequence 5178, Ap
					Sequence 2964, Ap
					Sequence 3159, Ap
					Sequence 18798, A
					Sequence 2, Appli
					Sequence 11, Appl
					Sequence 16, Appl
					Sequence 5080, Ap
					Sequence 6, Appli
					Sequence 1, Appli
					Sequence 1154, Ap
					Sequence 11, Appl
					Sequence 5756, Ap
					Sequence 8, Appli
					Sequence 6590, Ap
					Sequence 4347, Ap
					Sequence 1044, Ap
					Sequence 2, Appli
					Sequence 6, Appli
					Sequence 6, Appli
					Sequence 20284, A
					Sequence 3824, Ap

28	173.5	3.4	717	4	US-09-248-796A-18993	Sequence 18993, A
29	173.5	3.4	2954	4	US-09-150-867-1	Sequence 1, Appli
30	172	3.4	905	4	US-09-248-796A-16333	Sequence 16333, A
31	171	3.4	3241	4	US-09-841-786-1	Sequence 1, Appli
32	169.5	3.3	1010	3	US-09-134-001C-5178	Sequence 5178, Ap
33	169.5	3.3	1965	4	US-09-583-110-3829	Sequence 3829, Ap
34	169.5	3.3	1972	4	US-09-107-433-3251	Sequence 3251, Ap
35	169	3.3	1095	4	US-09-107-433A-3855	Sequence 3855, Ap
36	169	3.3	1874	4	US-09-331-403-2	Sequence 2, Appli
37	168.5	3.3	1027	3	US-08-446-137B-2	Sequence 2, Appli
38	168	3.3	2343	3	US-09-324-867-2	Sequence 2, Appli
39	167.5	3.3	584	4	US-09-949-016-10752	Sequence 10752, A
40	167.5	3.3	859	1	US-08-053-614-2	Sequence 2, Appli
41	167.5	3.3	859	2	US-08-316-397B-2	Sequence 2, Appli
42	167.5	3.3	859	1	US-09-034-306-2	Sequence 2, Appli
43	167.5	3.3	859	3	US-09-259-437-2	Sequence 2, Appli
44	167.5	3.3	859	5	PCT-US93-09782-2	Sequence 2, Appli
45	167.5	3.3	1181	1	US-08-053-614-4	Sequence 4, Appli
46	167.5	3.3	1181	1	US-08-316-397B-4	Sequence 4, Appli
47	167.5	3.3	1181	2	US-09-034-306-4	Sequence 4, Appli
48	167.5	3.3	1181	3	US-09-259-437-4	Sequence 4, Appli
49	167.5	3.3	1181	5	PCT-US93-09782-4	Sequence 4, Appli
50	167	3.3	872	3	US-08-851-843A-8	Sequence 8, Appli
51	167	3.3	872	3	US-08-851-843A-54	Sequence 54, Appl
52	167	3.3	872	3	US-08-974-549A-221	Sequence 221, App
53	167	3.3	872	3	US-08-854-050-8	Sequence 8, Appli
54	167	3.3	872	3	US-08-854-050-54	Sequence 54, Appl
55	167	3.3	872	3	US-09-430-323-8	Sequence 8, Appli
56	167	3.3	872	3	US-09-430-323-54	Sequence 54, Appl
57	167	3.3	872	4	US-09-402-181B-221	Sequence 221, App
58	167	3.3	872	4	US-09-721-456-221	Sequence 221, App
59	167	3.3	872	4	US-09-766-253-8	Sequence 8, Appli
60	167	3.3	872	4	US-09-766-253-54	Sequence 54, Appl
61	167	3.3	2733	4	US-09-949-016-11433	Sequence 11433, A
62	166.5	3.3	1078	4	US-09-583-110-4036	Sequence 4036, Ap
63	166.5	3.3	1080	4	US-09-107-433-4843	Sequence 4843, Ap
64	166	3.3	2366	1	US-08-480-604A-10	Sequence 10, Appl
65	166	3.3	2366	2	US-08-405-496A-10	Sequence 10, Appl
66	166	3.3	2366	3	US-08-915-136-10	Sequence 10, Appl
67	166	3.3	2366	3	US-08-957-310-10	Sequence 10, Appl
68	166	3.3	2366	4	US-10-011-366-10	Sequence 10, Appl
69	166	3.3	2366	4	US-09-084-517-10	Sequence 10, Appl
70	165.5	3.3	862	4	US-09-248-796A-16013	Sequence 16013, A
71	164.5	3.2	1979	4	US-09-949-016-6468	Sequence 6468, Ap
72	164.5	3.2	2047	4	US-09-949-016-7404	Sequence 7404, Ap
73	164	3.2	1279	4	US-09-710-279-3188	Sequence 3188, Ap
74	163.5	3.2	1534	4	US-09-543-681A-5182	Sequence 5182, Ap
75	163.5	3.2	2411	3	US-09-268-347-36	Sequence 36, Appl
76	162.5	3.2	1173	4	US-09-248-796A-19313	Sequence 19313, A
77	162	3.2	1964	2	US-08-790-912-3	Sequence 3, Appli
78	162	3.2	2052	2	US-08-790-912-2	Sequence 2, Appli
79	161.5	3.2	1282	4	US-09-134-000C-5785	Sequence 5785, Ap
80	161	3.2	886	4	US-08-956-171E-5235	Sequence 5235, Ap
81	161	3.2	886	4	US-08-781-986A-5235	Sequence 5235, Ap
82	161	3.2	990	2	US-08-645-193B-15	Sequence 15, Appl
83	161	3.2	1132	4	US-09-248-796A-15026	Sequence 15026, A
84	161	3.2	1307	4	US-09-949-016-7561	Sequence 7561, Ap
85	160.5	3.2	582	4	US-09-081-149-8	Sequence 8, Appli
86	160.5	3.2	956	4	US-09-107-532A-5007	Sequence 5007, Ap
87	160.5	3.2	1805	1	US-07-853-913-2	Sequence 2, Appli
88	160	3.1	1493	4	US-09-713-273A-20	Sequence 20, Appl
89	159	3.1	935	4	US-09-914-259-25	Sequence 25, Appl
90	159	3.1	1164	4	US-09-538-092-399	Sequence 399, App
91	159	3.1	1963	4	US-09-583-110-5243	Sequence 5243, Ap
92	159	3.1	1967	4	US-09-107-433-4883	Sequence 4883, Ap
93	158.5	3.1	1055	4	US-09-949-016-9776	Sequence 9776, Ap
94	158.5	3.1	1912	1	US-08-409-995-4	Sequence 4, Appli
95	158.5	3.1	1912	3	US-08-685-467-4	Sequence 4, Appli
96	158.5	3.1	2353	3	US-09-377-155-33	Sequence 33, Appl
97	158.5	3.1	2353	3	US-08-913-942-4	Sequence 4, Appli
98	158.5	3.1	2353	3	US-09-669-974-33	Sequence 33, Appl
99	158.5	3.1	2353	4	US-09-797-862-33	Sequence 33, Appl
100	158.5	3.1	2353	4	US-09-684-707-4	Sequence 4, Appli

101 158.5 3.1 2354 3 US-09-268-347-47 Sequence 47, Appl  
102 158.5 3.1 3289 2 US-08-477-451-2 Sequence 2, Appl  
103 158 3.1 1074 4 US-09-071-035-358 Sequence 358, App  
104 158 3.1 1074 4 US-09-071-035-394 Sequence 394, App  
105 158 3.1 1972 4 US-08-875-435B-3 Sequence 3, Appl  
106 158 3.1 2663 4 US-09-538-092-1252 Sequence 1252, Ap  
107 157.5 3.1 1001 4 US-09-248-786A-18658 Sequence 18658, A  
108 157.5 3.1 1164 3 US-08-923-992A-2 Sequence 2, Appl  
109 157 3.1 990 2 US-08-392-625-20 Sequence 20, Appl  
110 157 3.1 990 2 US-08-466-961A-20 Sequence 20, Appl

ALIGNMENTS

RESULT 1  
US-09-583-110-4263  
; Sequence 4263, Application US/09583110  
; Patent No. 6899703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 4263  
; LENGTH: 641  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-4263

Query Match 4.2%; Score 215.5; DB 4; Length 641;  
Best Local Similarity 23.2%; Pred. No. 8.7e-07;  
Matches 132; Conservative 72; Mismatches 178; Indels 187; Gaps 30;

QY 6 KTVALTTLTVSVV-----THNQVFS-----LVKEPILKQ 35  
DB 3 KTTILSLTAAVILAAVYVNEPILADTPSEVVKETKVGSIIOQNNIKYKVLTVEGNIRT 62  
QY 36 TOASSISGADY-AESSGK-----SKLXINETSGVDDTVTDLFS-----DK 76  
DB 63 VQVGVNGVTPVEAGQDGKPTIPTKTVGDKVFTVEVASQAFSYYPDETGRIVYYPSS 122  
QY 77 RTTPKIKDNLAKPREQELKAVTENTSEKQITSGSQLEQSK-----120  
DB 123 ITIPSSIKKIQKGFHSGKAKTII-----FDKGSQLEKIEDRAFDFSELEBELPAS 174  
QY 121 -----SLSLNK-TVPTSTNWEICDFTKGNLTVGLSKSGVEKLSQTDHLVPSQ 168  
DB 175 LEYIGTSAFSFSQKLGKLTFFSSSKLEL-----ISHEAFANLSNLEKLTLPKS 222  
QY 169 AAD-CTQLIQVA-----SPA-----FTPPKTAIAEYTSRAGENGISQLD 208  
DB 223 VKTLSGNLFRLTSLKHVDVEGNSFASVDGVLFSKO-KTQLIYPSQ--KNDESYKTP 279  
QY 209 VDGKEIINEGEVFNFSYLLKVKTIPTGYKHIGDADFVNDKNIAEVLNPSLETISYAF-A 267  
DB 280 KETKELASVSFNKNSY-LKKLELNEGLEKIGTGFADAKLBEISLPSNLETIERLAFVG 338  
QY 268 HLAHQIDLPDNLKAIGELAF-----FDNQITGKLSLPLQMLRAERAFKSNHIK 317  
DB 339 NLELKEILPQNVKRVFGKXVWNLGPKLKSLLTGNIN---SLPSPFLSGVLDSLKEIRK 395  
QY 318 -----TIEPRGNS--LKVIGEASFQ-DNDLSQLMLPDGLEKIEREAPT 357

DB 396 NKSTFSVKKDTFALPETVKFYVTSEHIKQVLKSNLSTSNDI-----IVEKVDNIKQETDV 451  
QY 358 GNPDDHYNNRVLWTK-----SGKNPISGLATENTYVNPDKSLW---QSPSEIDY 404  
DB 452 AKP-KKSNQGVGVWVKDKGLWYLYLNESSGSMATGMV-----KDKGLWYLYLNESSGM-A 502  
QY 405 TKWLEED--FTYQKNS-----VTGP-SNKGKL 427  
DB 503 TGMVKDKGLWYLYLNESSGSMATGMVKDKGL 531

RESULT 2  
US-09-107-433-2948  
; Sequence 2948, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-011  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 2948:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 641 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...641  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2948:  
US-09-107-433-2948

Query Match 4.2%; Score 215.5; DB 4; Length 641;  
Best Local Similarity 23.2%; Pred. No. 8.7e-07;  
Matches 132; Conservative 72; Mismatches 178; Indels 187; Gaps 30;

QY 6 KTVALTTLTVSVV-----THNQVFS-----LVKEPILKQ 35  
DB 3 KTTILSLTAAVILAAVYVNEPILADTPSEVVKETKVGSIIOQNNIKYKVLTVEGNIRT 62  
QY 36 TOASSISGADY-AESSGK-----SKLXINETSGVDDTVTDLFS-----DK 76

Db 63 VOVGNGVTPEFAGQDGRPTIPTKITVGDKVFTVTEVASQAFSPYDDETGRIVVYPS 122  
QY 77 RTTPEKIKONLAKPREQELKAVTENTSEKQITSGSQLEQSK-----120  
Db 123 ITIPSSIKIKOIKGFGSGAKTII-----FDKGSQLEKIEDRAFDFSELEBILPAS 174  
QY 121 -----SLSLNK-TVPSTSNWEICDFTKGNLTVGLSGVSKVGLSQTDLHLVPSQ 168  
Db 175 LEYIGTSAFSFSQKLKLTFFSSSKLEL-----ISHEAFANLSNLEKLTLPKS 222  
QY 169 AAD-GTQLQVA-----SFA-----FTPDKKTAEVTSRAGENGESOLD 208  
Db 223 VTKLSNLPRLTSLKXHVDEGNEFASVGVLFSD-KTQLIYPSQ--KNDESYPKTP 279  
QY 209 VDCKEIIINGEVFNYSLLKKVTIPTGKHIGODAFVNDKNIAEVNLPESLETISDYAF-A 267  
Db 280 KETKELASVFNKNSV-LKLELNEGLEKIGTFAPADAKLEISLPNSLETIERIAFYG 338  
QY 268 HIALKQIDLPNKAIGELAF-----FDNQITGKLSLPRQLMRLAERAFKSNHIK 317  
Db 339 NLELKEILPNDVNFNGKHMNGLPKLSLTIGNIN--SLPSFPLSGVLDLSLKEIHK 395  
QY 318 -----TIEFRGNS--LKVIGEASFQ-DNDLSQMLPDGLEKIESEAPT 357  
Db 396 NKSTEPSVKOTPAIPETVKFYVTSSEHIKDVLSNLSSTNDI-----IVEKVDNIKOETDV 451  
QY 358 GNPDDHYNRRVVLTK-----SGKNPGLATENTYVNPDKSLW---QESPEIDY 404  
Db 452 AKP-KKNSQGVGVKDKGLWYLNESGSMATGWV-----KDKGLWYLNESGSM-A 502  
QY 405 TKWLEED--FTYQKNS--VTGP-SNKGL 427  
Db 503 TGVVKDKGLWYLNESGSMATGWVKDKGL 531

## RESULT 3

US-09-949-016-6507  
; Sequence 6507, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6507  
; LENGTH: 3259  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-6507

Query Match 4.2%; Score 214.5; DB 4; Length 3259;  
Best Local Similarity 19.8%; Pred. No. 1.1e-05;  
Matches 212; Conservative 168; Mismatches 427; Indels 263; Gaps 45;  
QY 20 HNEVFSLVKPEILKOTQASSISGADYAESGSKLKI-----NETSGPVDPTVDL 72  
Db 233 HEDELLQLV-----TQAD-----VETENQKLRVLQRKLEHESLVGRAQVVDL 277  
QY 73 FSDKRTTPEKIDNLAGPREQELK-----AVTENTESEKQITSGSQLEQSKESL 126  
Db 278 LOBELTAABORNOILSQOQMEAEHNTLVNTEVEREESKILLERKMEVAERKLSF 336  
QY 127 TVPSTSNWEICDFTKGNLTVGL-----SKSGVEKLSQTDHLVLPQAAADGTQLIQV 178

Db 337 NLOEEMHLLLEQFEQAGQAQAELESRYSALEQKHKAEMEKTSHLSLQKTG---QELQS 393  
QY 179 ASFAPTPPKTAIAEYTSRAGENGEL-----SOLDVDGKEIINEGVFNYSLLKKVTIPT 233  
Db 394 ACDALKQNSKLLQDNEQAVQSAQTIQOLEDQLOQSKEL-----SQFLNRLPQQ 445  
QY 234 GYKHIGODAFVNDKNIAEVNLPESLETISDYAFHALLKQIDLPNKLKAIGELAFDNOI 293  
Db 446 -HETASQTSFPDVYN-----EGTQAVTEENIASLQKRVVL-----ENK 484  
QY 294 TKGLSLPQLMRL-AERAFKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKTE 352  
Db 485 GALLLSLIEBELKAENKLSQITLLEAQNRT---GEA---DREVSEISIVDIANKGS 537  
QY 353 SEAFVGNPGDDHYNRRVVLTKSGKNPGLATENTYVNPDKSLWQESPEIDYTKWLEEDF 412  
Db 538 SSA-----BESGD-----VLETFPSQKHKL-----SVLLEMKGAQE 573  
QY 413 TYQKNSVTG-----FSNKGLOKVKRNKNLE-----IPKQHNGVT 446  
Db 574 AFLKLQLOKRAEADHEVLQKEMKQMEGEGIAPIKMKVFLDTGQDPLMPNBSSESLP 633  
QY 447 ITEIGDNAPRVDFQNTLRKYDLBE--VKLPSTIRKIGAFQSNKLSKFSASDDLEI 504  
Db 634 AVE---KEQASTEHQSRSTSEISLNDAGVELKST-----KQDGDKLSLAVPDIGQC 681  
QY 505 KEGAFMNRRIETLELKDKLVITIGDAAFH-INHIYAIPLVSPVQEIQR-----SAFRQGA 558  
Db 682 HOELERLUKSOILEL-----ELNFHKAQEIYEKNDLDEKAKEISLNLQLEEFKKNAD 733  
QY 559 NN-----LIFMGSKVTKTGEMAFNSLEHLDL---SEOKLTEIPVQAFPS 601  
Db 734 NNSAFTALSEERDQL---SQVKELSMVTELRQVKQLEMLAEARERRLDYSESQTAH 790  
QY 602 DNALKEVL---LPASLKTIREEAFKKNHLKOLEVASALSHTAFNALDNDGDQEDPNK- 656  
Db 791 DNLLTEQIHSLSIEAKSKDVKIEVL-QNELDDVQLQFSEOSTLIRSL-----QSOLQNK 844  
QY 657 -----VVVTHNSVALADGEHFIVDPKLSSTIVDLKILKLEGLDYSTLR 704  
Db 845 SEVLEGAEVRHRISSKVEELSQALESQLEITKMDQLL-----LEK-----KRDVETLQ 893  
QY 705 OTTOFQDFMTTAGKALLSKNLQGEKQKQFLOEQAFFLGRVLDLKAIAEAKALVTKA 764  
Db 894 QIIEKDKQVT-----EISFSMTEKRVQVNEEKFSLG-VEI-KTLKQNLNLSRABE 943  
QY 765 TKNGQLLERSINKAVLAYNNSAIKKA---NVKRLKELDLTLGLVEGKPLAQATMVQGV 821  
Db 944 AKKEQVEEDNEVSSGLKQNYDEMSFAGQISKEELQHEFDLLKKNQKRLQALINRK 1003  
QY 822 YLLKTPPLPEYIIGLVNVPFKSGKLIYALDMSDT-IGEGQKDAYGNPILNVDENEGVH 880  
Db 1004 ELLQVRSRLEBELANLK---DESKK---EIPLSETERGEVEE-----DRENKEYS 1047  
QY 881 ALAVATLADYEGLDIKTILNSKLSQTSIR-----QVPTAAVHRAGI 922  
Db 1048 EKVTSKQOEIHYLKQIISEKEVELQHIRKDLKLEKLAEBEQFALVKQMOTLQDKNQ 1107  
QY 923 FOAIQNAAAAEQQLLPKPTTHSEKSSSES-ANSKDRGLQSNPKTNRGH 971  
Db 1108 IDLLQAEISENQAIQKLITSNTDASDGDGVALVKETVVISPPCTGSSEH 1157

## RESULT 4

US-09-134-000C-5178  
; Sequence 5178, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032



1423	Db	1423	QOLKEQPVAVDKETLKEQIAQ-ARGRKPEGYQFT---KETSKQLQEIQAQAAEIAVAKET	1478
727	Qy	727	LRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKA-----TKNQQLLERS---INKA	778
1479	Db	1479	ATKEEVSEALNALETMAAQLKEVPLVKNQDQLEVVKRAQQVTSSEGHQFTASSLSQELQKA	1538
779	Qy	779	VLAYNNSAIKK--ANVRKLEKELDLTLGLVEGKGLPLAQATWVGQVLLKTPPLPEYY--	834
1539	Db	1539	LLAAKNT--LKNPAANQKMWIDEAVAEVLSAIDG---LQEEVLVTDKKALEAMIAKAKAIKP	1594
835	Qy	835	IGLVNVPDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEHDEGHYHALAVATLADYEGL	893
1595	Db	1595	SAGKEFTSESARLATEAIDQAEGL-LADKNARQEQI-DIAEKN-----VKTALD--SL	1643
894	Qy	894	DIKTILNKLQSLSIROPVPTAAVHRAG--IFQAIQNAABE---AEOLLPKKPTHSEKS	947
1644	Db	1644	E-EQVLTQDTKTKLKLQKASTLPKPKAGKQFTKASQAEALABAIAKQALVEDPNATQEA	1702
948	Qy	948	SSSESANSK-----DRGLQSNPKTNRGRHSAI-----	974
1703	Db	1703	DKCLSILQSAIEAMAEEPISSNNTGNGNHSVSGTGVTSQKGTATGTTTTSST	1762
975	Qy	975	LPTGSKGSFYVIGILGY---TSVALLSLITAIKUKK	1007
1763	Db	1763	LPKANEVVSPISWSIGSFLLIVSIGLGLFKPKNKKEQ	1798
RESULT 5				
US-09-710-279-2964				
; Sequence 2964, Application US/09710279				
; Patent No. 6703492				
; GENERAL INFORMATION:				
; APPLICANT: KIMMERLY, WILLIAM JOHN				
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS				
; FILE REFERENCE: P0348008				
; CURRENT APPLICATION NUMBER: US/09710,279				
; CURRENT FILING DATE: 2000-11-09				
; PRIOR APPLICATION NUMBER: 60/164,258				
; PRIOR FILING DATE: 1999-11-09				
; NUMBER OF SEQ ID NOS: 4472				
; SOFTWARE: Patent in Ver. 2.1				
; SEQ ID NO 2964				
; LENGTH: 5024				
; TYPE: PRT				
; ORGANISM: Artificial Sequence				
; FEATURE:				
; OTHER INFORMATION: Description of Artificial Sequence: synthetic				
; OTHER INFORMATION: amino acid sequence				
; FEATURE:				
; NAME/KEY: MOD.RES				
; LOCATION: (5024)				
; OTHER INFORMATION: variable amino acid				
US-09-710-279-2964				
Query Match				
Best Local Similarity 4.0%; Score 205; DB 4; Length 5024;				
Matches 207; Conservative 165; Mismatches 389; Indels 298; Gaps 50;				
18	Qy	18	VTHNQEVSLVKEPIKQTOASSISGADYABSS--GKSKLKINETSQ-PVDDTVTLDFS	74
35	Db	35	ITMGQTIINKTADPVLDTKLVDNALISITKENALHGEQKLTTAKTEAINALNTLADL--	92
75	Qy	75	DKRTPPEKIKONLAKGPREQELKAVTENTSEKQITSGSQLEQSK-----ESL	122
93	Db	93	---NTPQK-----EAKTAINTAHTRTDVTAA-----EOSKANOINSAMHTLQNI	134
123	Qy	123	SLNKTVPSTSNWEICDFITKTNLTGLSKGVEKLSQTDHLVLPQAA-DGTOLIQVASF	181
135	Db	135	SDNESVTHESNYI-----NAEPEKQHAFTAEALNNAKEIVNEQQATLDANSINQKQA	186
182	Qy	182	AFTPDKTAIAFYTSRAGENG--EISQID-----VDGK	212

Db 187 ILTTKALDGEELRAKENADQIEINTLNQTDQARNSEKGLVNSQTRTEVASQALAK 246  
Qy 213 EIINEGEVNSYLLKKVTIPTGYKHIGQDAFVQNKNAEVLNIPESLETISDVAPAHALK 272  
Db 247 ELKNVMEQLNNLNGKNQMINSSKFINEA---NQOAYSNALASAEVLKNKS---QNP 299  
Qy 273 QIDLPNLKAIGELAFDQITQKLSLPQMLAERAFKSNHIKTIETFRGNSLKVIGEA 332  
Db 300 ELDKVTIQAINNINSAINNLNGEAKLTKA-----KEDAVASI----- 337  
Qy 333 SFQNDLSQLMLPDGLEKITESEAFNPGDDHYNNRVVLTKGKNPGLATENTVVPD 392  
Db 338 ----NNLSG--LTNEQTKENQAVGSQTRDQVAN--VLRDSKALDQS--MQLRDLVNQ 388  
Qy 393 KSLWQSSPEIDYTKMLEEDFTYKNSVTGFSNKGLOKVRKNKLEIPKQHGVTIPEIGD 452  
Db 389 NVIHSTF-----NYFNEDST-QKNTYDNDALNGSTYITCOHNSLNKSTIDOTISQI-- 439  
Qy 453 NAFRNVDFQNTLURKYDLEEVKLPTIRKIGAPAFQSNLKLKFEAS-----DDLEE- 503  
Db 440 NTAKN-DLHGAELQDRDKGTAN-----QBETGQLGYLNDPQKSAEESLVNGSNTRSEVEH 493  
Qy 504 IKEGAFMNRRIETLEKDKLVT-----IGDAAFINHIYAIVLPEB-----VQIEGR 550  
Db 494 LNEAKSLNAMK--QLRDKVAEKTNVQSSDYINDSTEH--QRGYDQALQEAENIINEIGN 550  
Qy 551 SAFRQNGANNLIFMGSKVTTLGEMAFLSNRLE--HLDLSEQKQ--LTEI-PVOAFSD--- 602  
Db 551 PTLNKSEIE-----QKQLQDQALQALQSHL--LEEAKNAITEINKLTALNDAQ 600  
Qy 603 ----NALKEVLPA-----SLKTIREAFKKNHLKOLEVASALSHTIAFNAL 644  
Db 601 QKAIENVQAQOTIPAVNQQLTLDRINTAMQALRDKVQGNVHQ-----QSNYFN-- 651  
Qy 645 DONDGEQDNKVVVKTNSVALADGEHFIVDPDKLSSTIVDLEKILKIEGLDYSTLR 704  
Db 652 EDEQPKHNDNSVQA-----GOTII---DKLQDPIMNKNIEQAINQIN----- 692  
Qy 705 QTTOTQFRDMWTAGKALLSKNSLRQEGKQFLQEAQFFLGRVLDLKAIAKAEKALVTKA 764  
Db 693 -TTOT-----ALSENKLHTDQESTNRQ-----IEGLSLNTAQINAEKDLVNQAK 737  
Qy 765 TKNQOLLERSINKAVLAYNNSAIKKANVRLEKELDLTLGLVEKGPLAQATWVQGVYLL 824  
Db 738 TR-----TDVAQKLATAKEINSAMSNLRDGIQNKEDIKRSS-----AYIN 777  
Qy 825 KTLPLPEYYIGL-----NVYFDKSGKLIYALDMSDTIGEGQDKAYGNPILND 873  
Db 778 ADPTKVTAYDQALQNAENIINATPNVELNKA-----TIEQAL-----SRVQQAQ 821  
Qy 874 EDNEGVALAVA-----TLADYEGLDIKTLNKLKLSLTISIRQVPTAAVHRAG 921  
Db 822 QDLDGVQQLANAKQQAQTQTVNGLSLNDQKRELNLINSANTR-TKVQELNKAATESNH 880  
Qy 922 IFQAIQAAAEABQLL-----PKPGTHSEKSSSES 952  
Db 881 AMBALRNSQNVQDVQKSNVYNEDDQEQHNDYNAVNEA 919

RESULT 6  
US-09-134-001C-3159  
; Sequence 3159, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynm Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3159  
; LENGTH: 10182  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3159

Query Match 4.0%; Score 205; DB 3; Length 10182;  
Best Local Similarity 19.1%; Pred. No. 0.00029;  
Matches 201; Conservative 178; Mismatches 447; Indels 224; Gaps 45;

Qy 47 YAESGKGLKINETSGPV--DDTVDFSDKRTTPEKIKDLAKGRPEQELKAVTENTE 104  
Db 7618 YNQAIVNAKAKNIINDQPTPMANDEIQSVLNEVKQT---KDNL-----HGDQKLANDKTD 7668  
Qy 105 SEKQITSSQSQESLSLNTKVTSTSNWEICDFTKGTNTLVGLSKSVGEKLSQTDHLV 164  
Db 7669 AQAATLNALNYLNAQGRNLETKVQNSNSRPEVKVQVLQANLQNDAMKKLDDALTGNDAIK 7728  
Qy 165 LPSQAADGTQLIQVASFAPTPDKTAIAEYTSRAGENGESOL-----DVGKELI 215  
Db 7729 QTSNVINEDTSQVNFDEYTDGKNIVAEQTPNPNMPTNINTIADIKTEAKNDLHGQVKL 7788  
Qy 216 NEGEVFNLYLLKKVT-IPTGYKHIGQDAFVQDNKNIAEVN-----LPESLETISDYAF 266  
Db 7789 KQAQOQSINTINQMTGLNQAQKEQLNQBETQOTRSEVHQVINKAQAALNDSMTLRQSIT 7848  
Qy 267 AHLAKQDLDLNLKAIGELAFDN-----QITKLSLPQMLAERAFKSNHIKTI 320  
Db 7849 DEHEVYKQTSNYIN-ETVGNQTAYNNAVDKVIQIINQTSNPTMNPLEVERA---TSNVK 7902  
Qy 321 PRGNSLKVIGEASFQNDLSQLMLP---DGLKIESEAFGTG-----NPGDDHYNNRV 369  
Db 7903 ISKDALH--GERELNDNKNKTFVNHLDNLNQAQKEALTHEIEQATIVSQVNNIYNKAK 7960  
Qy 370 VLWTKSGK-----NPSGLATENTYVNPD---KSLWQE-----SPEIDYTKW 407  
Db 7961 ALNDNMKKLKDIVAQDQNVQSNYINEDSTPQNNYNDINHAQSIIDQVANPTSHDE- 8019  
Qy 408 LEEDTYQKNSVTGFSNK-GLQKVRKNKLEIPKQHGVTITEIGDNAFRNVDFQNTLR 466  
Db 8020 TENAINNIKHAINALDGEHLKQQAENANL-----LINSLNDLNAQPDARNLVN 8070  
Qy 467 KYDLEEVKLPTIRKIGAPAFQSNLKSFEASDLEIEKEGAFMNRRIETLEKDKLVTI 526  
Db 8071 EAQTRF-----KV-----AEQLQSAQALNDAMKHLRNSIQNQ---SSVQESKYINA 8114  
Qy 527 GDA-AFHINHIYAI-----VLPEVQEIERSAFRQ-----NGANNLIFMGSKVKTIGEMA 575  
Db 8115 SDAKKEQYNH-AVREVENIINEQHPDLDEIKELIKQITDGVNQANDL----- 8159  
Qy 576 FLNSRLEHLDSQKQLTEIP---VQAFSONALKE-----VLLPAS 613  
Db 8160 ---NGVELLDADQNAHQSIPTLMLHNAQQAQNALNEKINNNAVTRTEVAALIGQAKLLDHA 8216  
Qy 614 LKTIREEAFKKNHLKOLEVASALSHTIAFNAL--DNDGDGEQPDN-----KVVKVTHNSY 666  
Db 8217 MENLEESIKDKEQVKQ-----SSNYNEDSDVQETDYNADVHTEILNQTVPNPTL 8266  
Qy 667 ALADGEHFIVDPDKLSSTIVDLEKILKIEGLD--YSTLRQTTQTFQFRDMTTAGKALLSK 724  
Db 8267 SIEDIEHAINEVNQAKQLRGKQKLYQITDLADKELSKLDDLTSSQSSISINQIYAKTR 8326  
Qy 725 SNLRGQ-EKQPLQBAQFPLGRVLDLKAIAKAEKALVTKATKNGQLERSINKAVLAYN 783  
Db 8327 TEVAQIAIEKAKSLNHAM-----KALNKVYKNADKVLDSRFINEDQPEKKAYQQAINHVD 8381  
Qy 784 NSAIKKAN-----VKLEKELDLTLGLVEKGKPLAQ-----ATWGVGVYLLKTPPLP 831  
Db 8382 SIIHRQTPNEMDPTVINSTIETLETAQNNLHGQDKLAHQDAANVING--LIHLNVQAR 8439  
Qy 832 EYIIGLVNVFQKSGKLIYALDMSDTIGEG-----QKDAYGNPILN-----VDEEDGNYHAL 882

```
Db 8440 EWMINTNTATREKAVKLDNAQALDKAMETLQOVVAHKNNILNDSKYLNEDSK-YQQQ 8498
Qy 883 AVATLADYEGDLTKILNSKLSQTSIRQVPTAAVYHRAGIFQAIQNAABAEQQLPKPGT 942
Db 8499 YDRVIADAEQL-----LNQTTN-----PTLEPKYKVDIVK--DNVLANEKILF----- 8538
Qy 943 HSEKSSSESANSKDRGLQSNPKTNRGHS 972
Db 8539 GAELKSYDKS-NANDEIKHMYLNNAQKQS 8567

RESULT 7
US-09-248-796A-18798
; Sequence 18798, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18798
; LENGTH: 924
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18798

Query Match 4.0%; Score 204.5; DB 4; Length 924;
Best Local Similarity 20.2%; Pred. No. 8.9e-06;
Matches 184, Conservative 163; Mismatches 351; Indels 213; Gaps 46;

Qy 178 VASFAPTDPKKTAI---ABYTSRA---GENGEISQLDVDGKEIINEGEVFNLSYLLK--KV 229
Db 20 VVTFFYQNTTRTFVDMATTTTSNAPANGENASITQIIVNHKNVSKADDLISFNISKQKL 79
Qy 230 TIPTGYKHGQDAFVDNKNIAEVLNPELSETISDYAFALHAKQJDLDPNLKAIGELAFF 289
Db 80 KVVNLDPYDGSFAFIDS-----SUKVSDP-----VKIPEPAN-SPVNSLAKT 122
Qy 290 DNQITGKLSPLQMLRL-----AERAFKSNHIKTIFRGNs---LKVIGEASFQDN 337
Db 123 NND-----TVTKQVRVPSTYRPTLSSSRHSPKNLSKRSRINSNSPDKMTLSKRVRLKNS 177
Qy 338 DLSQLMLPDGLEKISEAPTGNPGDDHYNRVVLTWKSGKNPSGLATENTYVNP----- 392
Db 178 PLSQ-----SVDSNPFI-----EKNSKIYL-----KN-LALRTSSTDTLTDSHH 217
Qy 393 -----KSLMQP-----SPEDYTKWLEDFYQKNSVTGFN-----KGLQKVRKNK 435
Db 218 DCCLQIETLQNEKRVHLHKYVDSIKDLQDFLAFQKQEAAGIINVLDNDAQLESLQK-Q 276
Qy 436 LEIPKQHNQVTTIEGDNAPRVNDFQNTLRKYDLEEVKLPSTIRKIGAFQSNLKSF 495
Db 277 IDLENTKNDQIKELVSEHLOIETIKSKLTKFKLVETQLLSK-HESHVQQTTELIAES 335
Qy 496 RASDLEETKEGAFMNNRIETLELKKDL-----VTIGDAAPHINHIYVLPESVQIG 549
Db 336 ESKQVVEELE-----NLKTHAKDSARILELETQLSDAA-----KEKS 374
Qy 550 RAFPQNGANNLI-FMGSKVKTGEMAFLSNLEH-----LD-LSEQKOLTIPIQA 599
Db 375 ESDYKLTDTSEIVNLDKSIETL--KANL-NKLEBEREIQNKLDQVSELKELKLVEE 431
Qy 600 FSDNAL-----KEVLLPASLKTIRREAFKKNHLKQLEVASALSIAFNALDND--- 648
Db 432 LSNLLKQQLHREKIELNEQLEKLGVSADKELQGSIEQ--SKIYINSKQNAKT 489
```

```
Qy 649 -----GDEQPDNKVVVTKTHNSYALADGEHFIVDPDKLSSTIVOLE---KILK--LI 695
Db 490 DRNWEIKYGTLEQNOQKLVBEH-----LANRQDQLEIKLVKSKTELEADNQAQKNQLT 544
Qy 696 E-----GLDYGTLL-RQTTOTQFR-----DMWTAGKALLSKNSLRQGEKQKPLQ 737
Db 545 EISKQFSTLGVSESELKKQLEBETQIKYDQAEIVHQELKAQIEKLLKESAGKQGLKELNQ 604
Qy 738 EAQFFLGRV--DLDKAIKAKAEKALVTKATKNGQL---LERSINKAVLAYNNSAIKXAN- 791
Db 605 SHEKFVNELKNEHEKQVKETKDQIIKEMEEKHQQAIEIEDSHNENIKENINNEHENKAKC 664
Qy 792 -VKLEKELDLTLGLVEGKGLAQTAVQGVYLLKTLPLPLPEYVIGLVNVPYDKSGKLIYA 850
Db 665 IIDLNEIEBELTSOL--KNAESENKTLQSLKT-----EYENELIAYKSKIDQL--- 711
Qy 851 LDMSDTTIGEGOKDAYGNPILNVNDENGYHALAVATLADYEGDLIKTILNSKLSQTSIR 910
Db 712 ---EKESAENLKE-YEAKLSMKFDESDLAIEKQLRKD-DGQDFENQI-EKLNQLYTEK 765
Qy 911 QVPTAAVYHRAGIFQAIQNAABAEQQLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGR 970
Db 766 DLQLEKSD-----BEIASIKKMEEBEDTKAKLEKSTTQNGMKLLDNLSTK-----K 813
Qy 971 HSAILPRTGSK 981
Db 814 HAAVVAELSGK 824

RESULT 8
US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCT
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-308-375-2

Query Match 3.9%; Score 199.5; DB 3; Length 2285;
Best Local Similarity 20.9%; Pred. No. 7.7e-05;
Matches 215; Conservative 155; Mismatches 356; Indels 303; Gaps 54;

Qy 81 EKIKNLAKGPREQELKAVTENTES-EKOITSGSQLEQSKESLSLN----- 125
Db 771 KKAKDDPEQS-QQTNVEAITTKDSTDKLIQQYKELQKVKESSRLTSDDEQEYLVQVTOOL 829
Qy 126 -KTPVSTSNWEICDFTKNGNTLVGLSKSGVEKLSQT-DHLVLPQAA-----DGTQ 174
Db 830 AQTFFAL-----VKGYDSQGNAILKTNKELEKAIENKEYLALKKQETRDSAKKTFEDASK 885
Qy 175 LIQVASFAPTPDKTATAIAYTSRAGENGEISQLDVGKEIINEGEVFNLSYLLKVVPTPTG 234
Db 886 EIKKSKDELKQYKQ--IADYNDKGRPKWDLTADDDYKVAADKAK--QSMLKASQSDIESG 941
Qy 235 YKHIGQDAFVDNKNIAEVLNPELSET-ISDYAFALHAKQJDLDPNLKAIGELAFFDQNI 293
Db 942 NAKVDSVLSTANAYSSIDISNTLKTSLSD-----VVKLNKLDLDDLP-BELEKPFSSSL 994
Qy 294 TGKLSLPRQLMRLAERAFKSNHIKTIFRGNLSKVIGEASFQDNDLSQMLPDGLEKIES 353
Db 995 -GKLEKQKQ-----KALDSGDEKAFD---NAKK-----DLQSL----- 1024
```

```
QY 354 EAFTGNPGDDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEDEFT 413
Db 1025 -----ETYSKSDSS-----IDVFKMSFD--K 1043
QY 414 YQKSVTGSNKGLOKVKRNKNLEIPKQHNGVTITTEIGNAFNVDVFNQKTKRYDLEEV 473
Db 1044 AQKNIKDG--DKSLSSVKSEVG-----DLGETLAEAGNEA---EDFGKK-----LKEA 1086
QY 474 KLPSTIRKIGAPAFQSNLKSPEASDLLEIEKEGAFMNNRIETLEKDKLVITIGDAAFHI 533
Db 1087 LDANSVDDIKAAIKEMSDAMQF--DSVDVLNGDIFNN-----TKQVAPLND----- 1132
QY 534 NHIYAIVLPESVOEIGRSFRQNGANNLIPMGSKVKTGEMAFNLSRHLHDLSEKQLT 593
Db 1133 -----LLEKMAE--GKS--ISANEANTLI---QKDELAQAISIENGVVVKINRDEVIKQR 1180
QY 594 EIPVQAFSD-----NALK-----EVLIPASLKTIREE----- 620
Db 1181 KVKLDAYNDMVVTSYNNKLMKTEVNNAIKTLNADTLRIDSLKLRKERKLDMSAEALSDELV 1240
QY 621 -----AFKNHKLKOLE-----VASAL--SHIAFNALDDNDGDEQFD 654
Db 1241 KSNINNVADAKKELKLEKMLQPGGYSNSQIEAMQSVKSALESYI--SASEATSTQEMN 1298
QY 655 NKVVVTHHNSYALADGEHFIIVDPDKLSSTIVD-----LEKILKLIIEGL-----DYSTLR 704
Db 1299 KQALVEAGTSLNWTDOQEKANEETKTSYMYVDVKYKEALEKVNAEIDKYNKQVNDYPKYS 1358
QY 705 QTTQTQFRD--MTTAGKALLSKNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAFALVTKK 763
Db 1359 Q-----KYRDAIKKEIKALQOKKKLMQ--EQAKLLKD--QIKSGNI-----TQYGIWTS 1404
QY 764 ATKNG--QLLERSINKAVLAYNNSAIKKANVKRLEKLELDLTLGLVEGK--GPLAQATMVOG 820
Db 1405 TSGGTSPSTGGSGYSKYSYNSAASKYNV-----DPALIAAVIOQESGFNAKARSGV 1459
QY 821 VYLLKTPLPPEYIYGLNVYFDSKGLIYALDMSDTIGEGOK--DAYGNPILNVDEBNEY 879
Db 1460 AMGLMQLMPATAKSLGVNNAYDP-----YQNVMGGTKYLAQOLEKEGG--NVEKALAY 1511
QY 880 HALAVATLADYEGLD-----IKTIL---NSKLSQLTSIROVPTAAYHRAGIFOAION 928
Db 1512 NA--GPGNVIKYGGIPPEKFTQNVVKIMANYSKLSATS-----SIASY----- 1556
QY 929 AARAEQLPKPQTHSEKSSSESANSKDGLOSNPKTNRGRSAILPRTSGK--SFVY 986
Db 1557 -----TNSAFRVSSKYQQESGLRSP--HKGTDFAAKAGTATKSLQSGV 1601
QY 987 GILGYTSA 995
Db 1602 QIANGSKTA 1610
```

```
RESULT 9
US-09-932-183A-2 ;
; Sequence 2, Application US/09932183A
; Patent No. 6833265
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
```

```
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-932-183A-2
```

Query Match 3.9%; Score 199.5; DB 4; Length 2285;  
Best Local Similarity 20.9%; Pred. No. 7.7e-05;  
Matches 215; Conservative 155; Mismatches 356; Indels 303; Gaps 54;

```
QY 81 EKIKNLAKPPEQELKAVTENTES--EKQITSGSQLEQSKESISLN----- 125
Db 771 KKAODDFEOS--QOTNVEAITTKDSTDKLIQYKSLQKVKRSRSLTSDEBEYLVQVTOQL 829
QY 126 -KTVPSTSNWEICDFTTKGNTLVGLSKSGVEKLSQT--DHLVLPQAAA-----DGTQ 174
Db 830 AQTFPAL-----VKGYDSQGNALTKNKELEKAIENTKEYLALKKQTRDSAKTTFEDASK 885
QY 175 LIQVASFAPTPDKTAIAEYTSRAGENGEISOLDVDGKEIINEGEVFNYSLLKCVTIPGT 234
Db 886 EIKKSKDELKQYKQ--IADYNDKGRPKMDLIAADDDYKVAADKAK--QSMLKAQSDIESG 941
QY 235 YKHIGQDAFVDNKNIAEVLNLPESLET--ISDYAFALHALKQIDLPNLKAIIGELAFEDNOI 293
Db 942 NAKVKDSVLSTIANAYSSIDISNTLKTSTSD-----VNVKLNKDDOLDP--EELEKPS 994
QY 294 TGKLSLPRQLMLAERAFKSNHIKTEIFRGNLSKVIGEASFQDNDLSQLMLPDGLEKIES 353
Db 995 -GKLEQKQV-----KALDSGDEKAFD---NAKK-----DLQSL----- 1024
QY 354 EAFTGNPGDDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEDEFT 413
Db 1025 -----ETYSKSDSS-----IDVFKMSFD--K 1043
QY 414 YQKSVTGSNKGLOKVKRNKNLEIPKQHNGVTITTEIGNAFNVDVFNQKTKRYDLEEV 473
Db 1044 AQKNIKDG--DKSLSSVKSEVG-----DLGETLAEAGNEA---EDFGKK-----LKEA 1086
QY 474 KLPSTIRKIGAPAFQSNLKSPEASDLLEIEKEGAFMNNRIETLEKDKLVITIGDAAFHI 533
Db 1087 LDANSVDDIKAAIKEMSDAMQF--DSVDVLNGDIFNN-----TKQVAPLND----- 1132
QY 534 NHIYAIVLPESVOEIGRSFRQNGANNLIPMGSKVKTGEMAFNLSRHLHDLSEKQLT 593
Db 1133 -----LLEKMAE--GKS--ISANEANTLI---QKDELAQAISIENGVVVKINRDEVIKQR 1180
QY 594 EIPVQAFSD-----NALK-----EVLIPASLKTIREE----- 620
Db 1181 KVKLDAYNDMVVTSYNNKLMKTEVNNAIKTLNADTLRIDSLKLRKERKLDMSAEALSDELV 1240
QY 621 -----AFKNHKLKOLE-----VASAL--SHIAFNALDDNDGDEQFD 654
Db 1241 KSNINNVADAKKELKLEKMLQPGGYSNSQIEAMQSVKSALESYI--SASEATSTQEMN 1298
QY 655 NKVVVTHHNSYALADGEHFIIVDPDKLSSTIVD-----LEKILKLIIEGL-----DYSTLR 704
Db 1299 KQALVEAGTSLNWTDOQEKANEETKTSYMYVDVKYKEALEKVNAEIDKYNKQVNDYPKYS 1358
QY 705 QTTQTQFRD--MTTAGKALLSKNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAFALVTKK 763
Db 1359 Q-----KYRDAIKKEIKALQOKKKLMQ--EQAKLLKD--QIKSGNI-----TQYGIWTS 1404
QY 764 ATKNG--QLLERSINKAVLAYNNSAIKKANVKRLEKLELDLTLGLVEGK--GPLAQATMVOG 820
Db 1405 TSGGTSPSTGGSGYSKYSYNSAASKYNV-----DPALIAAVIOQESGFNAKARSGV 1459
QY 821 VYLLKTPLPPEYIYGLNVYFDSKGLIYALDMSDTIGEGOK--DAYGNPILNVDEBNEY 879
Db 1460 AMGLMQLMPATAKSLGVNNAYDP-----YQNVMGGTKYLAQOLEKEGG--NVEKALAY 1511
QY 880 HALAVATLADYEGLD-----IKTIL---NSKLSQLTSIROVPTAAYHRAGIFOAION 928
Db 1512 NA--GPGNVIKYGGIPPEKFTQNVVKIMANYSKLSATS-----SIASY----- 1556
```

```
QY 929 AAAEAEQLPKPGTHSEKSSSESANSXDRGLQSNPKTNRGHRSAILPRTGSKG--SFVY 986
Db 1557 -----TNSAFRVSSKYGQSGSLRSSP--HKGTDFAAKAGTAKLSQSGKV 1601
QY 987 GILGYTSA 995
Db 1602 QIAGYSKTA 1610

RESULT 10
PCT-US93-07261-11
; Sequence 11, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESS: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1488 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-07261-11

Query Match 3.9%; Score 196.5; DB 5; Length 1588;
Best Local Similarity 21.1%; Pred. No. 7.3e-05;
Matches 209; Conservative 144; Mismatches 372; Indels 267; Gaps 53;

QY 46 DYAGSS-GSKLKNINETSQVDD-TVTDLFSDKRTT-----PEKIKONLAKGPREQEL 96
Db 611 EYDQSELAKGDIITKPHSVDEYDQTELAKGKVTNKPENLEHYNETDLAKG----- 664
QY 97 KAVTENTESEKQITSGSQLESLSLNKTVPTSTNWEICDFTKGNLTVLGSKSGVEK 156
Db 665 KEVTKNPHSVDEYDQSELAKGKD--ITNKPESVDEYDQTE-LAKGKVTNKARENLEE 721
QY 157 LSQTHLVLPQAADGTQLIVASFAFTPKKTATA---EYTSRAGEN-GEISQLDV-DG 211
Db 722 YNETD-----LAGKVTNKARENLEHYNETDLAKGKVTNKARENLEHYNETDLAKG 774
QY 212 KEIINEG---BVFNSYLLKKVITPTGYKHIGQDAFVDNKNIAEVLNPLESLETISDYAPA 267
Db 775 KEVTKNARENLEHYNETDLAK-----GKE--VTNK--AHENLEHYNET--DLAKG 818
QY 268 HUALKQI--DLPDNLKAIIGELAFF-DNQITGKLSLPQLMLABRAFKSNHKTTFEFGN 324
Db 819 ----KEVTKNARENLEHYNETDLAKGKVTNK-----ARENLEHYNETDLAKGK 863
QY 325 SLKVIAGEASFQDNDLSQLMLPGLEKIESEAFGTGPGDDHYNRVVLMTKSGKNPSGLAT 384
```

```
Db 864 --EVTNKARENLEHYNETDLAKGKE--VTNKA---RENLEEYNETDL---AGKEVTNKA 914
QY 385 ENTYNPDKSLWQSSPRIDYTKMLE-----EDFTYQKNSVTGFSNKGLOKQVKR 432
Db 915 ENL-----EYNETDLAKGKVTNKARENLEHYEYEEKDYMKNN--ELQNKGSGLKE 963
QY 433 NKNLEIPKQHNGVITIIGDNAPFNVDPPONKTLRKLYDLEEVKLPSTTRKIGAPAFQSNL 492
Db 964 NAELKNKELRN-----KSGDGLKENAELKNKELNKGSDGLK-----ENAE 1005
QY 493 KSFE-ASDDLEIEKEGAFMNNRIETLEKDKLVITGDAAFHINHIYAVLPESVQIEGRS 551
Db 1006 KKNELNKGSEGLKENAELKNK-----ELRNK-----GSEGLKEN-----AELKNK 1046
QY 552 AFRQNGANNLIFMGSKVTKLGEMAPLSNRLEHDLSEKQQLTEIPVQAFSDNA-LK-EVL 609
Db 1047 ELRNKGSSEG-----LKENAELKN-----KELQNKSGSEGLKENAELKNKEL 1086
QY 610 LPASLKTIREAPKKNHLKQLEVASALSIAFNALDDNDGDDEQFDNKVVVTKTHNSYALA 669
Db 1087 QNKSGSEGLKENAEQKN--KELQ-----NKGSEGLKENAELKNKELRNKGS 1129
QY 670 DGEHFIVDPDKLSSTIVDLKILKILBGLDYSTLRQTTOTOTOFDMTTAG-----KALLSK 724
Db 1130 DGL-----KENAELKNKELRNKGSQ-----LKENAELKNKELRNKGSEGLKENVYTN 1177
QY 725 SNLRQGEKQKFLQBAQFPFLGRVDLDKAIKAKAKAL-----VTKKATKNQQLLERSI----- 775
Db 1178 NDKNNDIQ-----NKLDSNKMKNKELNKLNDISNKMKNKELLNKLDSNEDMK 1226
QY 776 NKAIVLYNNSAIKAVNVRLEKELDLLTGL--VEGKGPLAQATWVGQVYLKTLPLPLPEY 833
Db 1227 NKEILL--NKDIRNKDLKISGNMBEQNTGLKNTSPSKQ-----QNTGLKNTP----- 1270
QY 834 YIGLVYFDSKGLIYALDMSDTTIGEGQKDAYGNPILNVDSNDEGYHALAVATLADYBGL 893
Db 1271 ----NERQNTG-----LKNTPSEGOQNTGLKNTPSEGOQNTG----- 1304
QY 894 DIKTLNSKLSQLTISRQVPTAAYHRAGIFQAIQNAABAEQLLPKPTHSEKSSSSSA 953
Db 1305 -LKNTPNER-QQNTGLKNTPSEGOQNTG-----LKNTPIEGOQNTGLKNTPSEGOQNTGLK 1358
QY 954 NSKDRGLQS---NPKTNRGHRSAILPRTGSKG 982
Db 1359 NAANKGOQNTGLKNAANKGOQNTGLKNTPSTGK 1390

RESULT 11
PCT-US93-07261-16
; Sequence 16, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESS: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
```

NAME: Blásdale, John H. C.  
REGISTRATION NUMBER: 31,895  
REFERENCE/DOCKET NUMBER: DX0288K  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-822-7398  
TELEFAX: 201-822-7039  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1663 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
STRAIN: Malayan Camp  
PCT-US93-07261-16

Query Match	3.9%;	Score 196.5;	DB 5;	Length 1663;
Best Local Similarity	21.1%;	Pred. No. 7.8e-05;		
Matches 209;	Conservative 144;	Mismatches 372;	Indels 267;	Gaps 53;
Qy	46	DYABSS--GSKSLKINETSGPDD--TVTDLFSDKRTT-----PEKIDKNLAKGPRQEL	96	
Db	611	YDQSELAKGDIYTKPHESVDEYDQTELAKGKEVTNKPHEENLEENETDLAKG-----	664	
Qy	97	KAVTENTSEKQITSGSQLESLSLNTKVPSTSNWEICDFITTKGNTLVLGSKSGVEK	156	
Db	665	KEVTNKPHEVDEYDQSELAKGKD--ITNKPHEVDEYDQTE--LAKGKEVTNKARENLEE	721	
Qy	157	LSQTDHLVPSQAADGQTLQVASFATPDPKTAIA---EYTSAGEN--GEISOLDV--DG	211	
Db	722	YNETD-----LAKGKEVTNKARENLEENETDLAKGKEVTNKARENLEENETDLAKG	774	
Qy	212	KEINEG----EVNSYLLKKVITPTGYKHIGQDAFVDKNKIAEVLNPESLETISDYAPA	267	
Db	775	KEVTNKAHENLEENETDLAK-----GKE--VTNK--AHENLEENET--DLAKG	818	
Qy	268	HLAKQI--DLPDMLKAIGELAFF-DNOITGKLSLPQLMLRAEPKSNHIIKTIEPRGN	324	
Db	819	----KEVTNKARENLEENETDLAKGKEVTNK-----ARENLEENETDLAKGK	863	
Qy	325	SLKVIIGASPDNDLSQMLPDGLEKTESAFTGPGDGHVNNRVVLTKSGKNPGLAT	384	
Db	864	--EVTNKARENLEENETDLAKGKE--VTNKA---RENLEENETDL---AKGKEVTNKAR	914	
Qy	385	ENTYVNPDKSLWQSPSIDYTKWLE-----EDFTYQKNSVTGFSNKGLOKVKR	432	
Db	915	ENL-----EYNETDLAKGKEVTNKARENLEENEEYEEKDMKN--ELQNKSGDLKE	963	
Qy	433	NKMLEIPKHNGVTIITIGNAFRNVDFONKTLKYDLEEVKVLPSSTRKTKGAFAPQSNL	492	
Db	964	NABLKNKELRN-----KGSGLKENABLKNKELRNKSGSDGLK-----ENAE	1005	
Qy	493	KSPE-ASDLDLEEKEGAFMNNRIETLEKOKLVITIGDAAFHINHIYAILVPESVQETGRS	551	
Db	1006	KNKELRNKSGEGLKENABLKNK-----ELRNK-----GSEGLKEN-----AELKNK	1046	
Qy	552	AFRONGANNLIFMGSKVKTILGEMAFPLSNRLEHLDLSEBQKQJTEIPVQAFSDNA--LK--EVL	609	
Db	1047	ELRNKSGEG-----LKENAELKN-----KELQNKSGEGLKENAELKNKEL	1086	
Qy	610	LPASLKTIRBEAPKKNHLKQLEVASALSHTAFNALDNDGDQPDNKVVVKTTHNSVALA	669	
Db	1087	QNKSGEGLKENAEOKN--KELQ-----NKGSEGLKENAELKNKELRNKGS	1129	
Qy	670	DGEHFIYDPOKLSSTIVLEKILKLIBGLDYSTLRQTQTQFRDMTAG-----KALLSK	724	
Db	1130	DGL-----KENAELKNKELRNKSGD-----LKENAELKNKELRNKSGEGLKENVYTN	1177	
Qy	725	SNLRQGEKQKPLQEAQFPLGRVLDKAIKAAEKAL-----VTKKATQNGQLLERSI--	775	
Db	1178	NDLKNDNDIQ-----NKLDSNKDMKNKELLNKDIINSKOMKNKELANKOLDSNMDK	1226	

RESULT 12  
 US-09-134-001C-5080  
 Sequence 5080, Application US/09134001C  
 Patent No. 6380370  
 GENERAL INFORMATION:  
 APPLICANT: Lynn Doucette-Stamm et al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: GFC-007  
 CURRENT APPLICATION NUMBER: US/09/134,001C  
 CURRENT FILING DATE: 1998-08-13  
 PRIOR APPLICATION NUMBER: US 60/064,964  
 PRIOR FILING DATE: 1997-11-08  
 PRIOR APPLICATION NUMBER: US 60/055,779  
 PRIOR FILING DATE: 1997-08-14  
 NUMBER OF SEQ ID NOS: 5674  
 SEQ ID NO 5080  
 LENGTH: 3696  
 TYPE: PRT  
 ORGANISM: staphylococcus epidermidis  
 US-09-134-001C-5080

Query Match	3.8%;	Score 195;	DB 3;	Length 3696;
Best Local Similarity	18.8%;	Pred. No. 0.00033;		
Matches 220;	Conservative 164;	Mismatches 478;	Indels 310;	Gaps 45;
QY	10	LTLTTSVWTHNQEVSLVKPELILKOTQASSISIGADYAESGSKSLKINETS	GPVDDTV	69
DB	1082	ITHTNVVQKQPQARQALIAKTWEKQSAINSNDEGT--IEEKAKIQSLNDAKNLADEQI		1139
QY	70	TDLFSDKR-----TTPEKIKDNLAK-----GPREQELKAVTENTESEK	108	
DB	1140	TOAASNQVNDNALNIGISNISKIQTFTKKQARDQVQKQFQBEAELNSTPHATQDEKQ		1199
QY	109	ITSGSQLEQSGESLSLNKTVSPSTNSWEICDFITKG-----		143
DB	1200	-DALTRLTOAKET--ALNDINQAQONTQVDTALTSGIQNIQNTQVNVKQKEAKTTINDIV		1257
QY	144	-----NTLVGLSKSGVEKLSQTDHLLVLPQO---AADGTQLIQV		178
DB	1258	QQHKQSIQNDNDATTBEEKVANNLVNASQV--ISKIDNATTNNQIDGIVSDGRQIN-		1314
QY	179	ASFAFTP-----DKTAIAEYTSRAGE-----NGEISQLDVGDKIIL-		215
DB	1315	---AITPDTISIKRNAKNDIDIKAAADKKIKIORDINATDDEBIEANRKIEEAKTEAKDNIQ		1371
QY	216	-----HEGEVFNISLLKKVITPTCYKHIQODAFVDNK-----NIAEVLPESL		258
DB	1372	RNSTRQVNEAKTNGINKINIENTPATTVKSEARQA-QVQNKANQINHIQNTPOATNEEQ		1430
QY	259	ETISDVAFAPHLAKIQDLPDLNKAIGELAPFDDNOITGKLSLPQLQM--RLAEAFKSNHI		316
DB	1431	EALNRYS--AELARVQAIINAEHTTQGVTKTIKDDAITSLSRINQAQVKEKESARNAIEQAT		1489
QY	317	KTIEFRGNSLUKVIGEASFQDNDSLQMLPDGLKEIKISEAFTGNPGDDHYNRVVLWTKSG		376



```
Db 1490 QQTQFINNDNATDEEKEVANNLVATKQKSLDINSLS-----SNNDEVAKVAG 1540
Qy 377 KY-----PSGLATENTYVNPDKSLWQSPSPEDYTKWLEEDPTYQKNSVTFPSNKGLOKV 430
Db 1541 INEIANVLPAVAKSKAKDIDQKLAQTNQIQ-----THQ-TATTEKEAAIQOLA 1590
Qy 431 KENKN---LEIPKQNGVYTIIEIGDNAPFNVDQKTLRKDYLEEVKLPSTIRKIGA---484
Db 1591 NOKSNEARTAIQENSHNGVAQKNGIHBI-----ELVMPDAHKSDAKOS 1637
Qy 485 ----PAFOSNNLKSFEASDLEIEIKEGAFMNNRIETLSLKDKLVIGDAAFHINHIYAIV 540
Db 1638 IDNKYEQSNTINT--TPDATDEEQKAL-----DKLKTAKDAGY--NKVDQAQ 1682
Qy 541 LPESVOETGRSFRONGANNLIFMGSKVKTIEMAFPLNRLHLDSQKQLTEIPVQAP 600
Db 1683 TNQVSDAKTEAI--DTITNI--QANVAKPSARVELDSKPEDL-----KRQINATP---1730
Qy 601 SNALKEVLLPASLTIREEAFKKNHLKQLEVASALSHTAFNALDDNGDQFQDNKVUVK 660
Db 1731 --NATEE-----KODAIQRLNGKQDEVKNLIN-----QQRDNEV---1764
Qy 661 THNSYALADGHEFVDPDKLSSTIVDL-----EKILKLIRGL 698
Db 1765 EQHKNGIGLELETHANPTKSDALQELQTKFISQTELINNNKDATNEEKEAKELLEIS 1824
Qy 699 DYSTLRQTTQTO-----PRDMTAGALLSKSNLRQGEKQKPLQBAQPFQGR- 745
Db 1825 KNTKTINQAQTNQVNDKNGMNEIATIPATTIKTDAKTAIDKKAEOQVTTINGNN 1884
Qy 746 --VDLDKAIK--AEKALVTKKATNGQLLESINKA---VLANN---SAIKKANVKR 794
Db 1885 DATDEKAEARKLVEKAKIEAKSNTNSDTERVNGAKTNGLEKINNIPSTQTKNAQ 1944
Qy 795 -----LEKELDLT-----GLVEGKGPLAQATMVGVYLLKPLPL 830
Db 1945 BINDKAQEQLIQINNTPDATDEEKEQENATRVNAGLAQAIQINNASHSTQEVNESKTSIA 2004
Qy 831 PYYIGLVNYPFKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYHALAVATLAD- 889
Db 2005 TIKSVQPNVI-----KPEATINSLTQEAANNQKTLIGDNGNATDDEKAAKQLVTOKLEQ 2059
Qy 890 ----VEGLDIKTILSKLSQTSIRQVPTAAVHRAGIFQAIONAA--AEAEQLLPKPGTH 943
Db 2060 IQKHESQDNQDQNVKQAQITAIKLINANAHKQDQAINILTLAESKKSDIRANQDATT 2119
Qy 944 SEKSSSESANSKDRGLQSNPKTNRGHSAIL 975
Db 2120 BEKTAIQSID--DTLAQARNNINGANTNALV 2149
```

## RESULT 13

```
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022e1 Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6
```

```
Query Match 3.8%; Score 192.5; DB 1; Length 2482;
Best Local Similarity 20.2%; Pred. No. 0.00027;
Matches 222; Conservative 178; Mismatches 400; Indels 301; Gaps 52;

Qy 6 KTVALTITTVSVVTHNQEVFSLVKPEILKQTOASSISGADYAESGSKSLKINETSGPV 65
Db 1341 KTTALDQSEKQKTEKTOEL-----ESHQSECLHCIQVAEAEVKEKTELLQT---L 1387
Qy 66 DDTVTDLPFSDKRTTPEKI---KDNLAGPPEQELKAVTENTESEKQITSGSQLESKES 121
Db 1388 SSDVSELLKDTHLQEKLOSLEKSOALSUTKCELENQIAQINKEKELL-----VKES 1440
Qy 122 LSLNKTVPSTSNWICD-----FITKGNLTVLGS-----KSGVEKLS---QTD 161
Db 1441 ESLQARL--SESDYEKLVNSKALEALVEKGEFALRLSSQTEVHQLRGIEKLRVRIEAD 1499
Qy 162 -----HLV--LPSQAADGTQLIQVASFAPTDKKTATAEYTSRAGENGESQLVDVG---211
Db 1500 EKKQLHIAEKLKERENDSL-----KDKVENLERELQMEENQELVILDAENSKA 1550
Qy 212 -----KEIINEGEVFNYSLLKKVTIPTGYKHI-----GQ-----DAF 243
Db 1551 EVELTKQTQIEEMARSLKVFE---IDLVTLRSEKENLTQIOEQKQQLSELQKLSLSPKL 1607
Qy 244 VDNKNIAEVNLPLESETISDYAFAHKALKQIDLPNLKAIGELAFFDQITGKLSLPRQL 303
Db 1608 LEEKEQAEIQIKERSKTAVEM-----LQNQLKEL-----NEVAALCGDQEI 1649
Qy 304 MRLAERAFKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQMLPDLGLEKIESEAFNGPDD 363
Db 1650 MKATEQSLODPPIEEHQRLR--NSIEKL--RARLEADEKKQLCVLQQLKSEHHA-----D 1700
Qy 364 HYNRRV-----VLWTKSGKNPGLATENTYVNPDKSLWQSPSPEDYTKWLEEDPTYQKN 417
Db 1701 LLKGRVENLERELSEIARTNQEHAALEANS-----KGEVETLKAKIEGTQ---1746
Qy 418 SVTGFSNKGLO-----KVRKNKNLEIPKQNGVYTIIEIGDNAPFNVDQKTLRKDYDL 470
Db 1747 ----SLRGLDLVVTVIRSEKENLTNLQEQERISELEIINSSPENI--LQEKQEKVQM 1800
Qy 471 EEVKLPSTIRKIGAFQFOSNNLKSFEASDLDLEEK--EGAFMNNRIETLEKDKLVIGDA 529
Db 1801 KEKS--STAMEMLOTQLKELNVAALHNDQEAQKAKEQNLSSQVECELEBAKQLQGLD 1858
Qy 530 AFHINHIAIVLPESV-----QEIGKSAFRONGANNLIFMGSKVKTLGEMAFPLNRLHL 585
Db 1859 EAKNYY----IVLQSVNGLIQEV-----EDGKQKL-----EKDEISRLKNQIQ---1900
Qy 586 LSEQKQLTEIPVQAFSDNAL--KEVLLPASLKTIRREAFKKNHLKQLEVASALSHTAFNAL 644
Db 1901 --DQEQVLVSKLSQVEGEHQLWKQNQL--ELRNLTVLELFQKIQVLQSKNAS-----L 1947
```

QY 645 DNDGDEQPDNKKVVKTHNSVALADGHEFIVDPDKLSTIVDLKILKIEGLDYSLR 704  
Db 1948 QD-----TLEVLQSSYKNLENELELTMDKMSF-----VEKVNKM-----TAK 1985  
QY 705 QT-TOTQPRDMMTAGKALLSKNSLRQGEKQKFLQEAQFPLGRVLDLKAIAKAELVTK 763  
Db 1986 ETELQREHMHM--AQKTAELQBEL-SGEKNRLAGELQLLLEI-----K 2026  
QY 764 ATKNGQLLERSINKAVLAYNNSAIKKA-----NVKRLKEKELDILTLGLVKGKGLAQATWVQ 819  
Db 2027 SSKD-QLKE-----LTLENSELKSLDCMHKQVKEGKGVREIEAEVQLRHEAEK 2078  
QY 820 GYVLLKTPPLPYEYIGLVNVDK-----SGKLIYALDMSDTIG 858  
Db 2079 QALLLDTN---KOYEVEIQTREKLTSECECLSSQKLEIDLKSSKEELNNSLKATQTIL 2135  
QY 859 EQOKDA-----YGNPILNVDDNNGYHALAVATLADYEGLDIKTILNSKLSOL----- 906  
Db 2136 BELKTKMDNLKYVNLKKNENRAGQKMKLLKSCQLE--BEKEILOKELSOLOAAQEK 2193  
QY 907 -----TSIRQVPTAAVHRAGIFQAIQNAAAAEAEQLPKPG-----THSEKSSSESAN 954  
Db 2194 QKTGVTMDTKVDELTEIKE---LAKETLEKTKEADEVLDKYCSLLISHEKLEKAKEMLE 2250  
QY 955 SKDRGLQSNPKTNRGRHSAIL 975  
Db 2251 TQVAHLCSQQSDSRGSPLL 2271

RESULT 14

US-08-353-700-1  
; Sequence 1, Application US/08353700  
; Patent No. 5599919  
; GENERAL INFORMATION:  
; APPLICANT: YEN, TIMOTHY J.  
; APPLICANT: RATTNER, JEROME B.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
; STREET: 1601 MARKET STREET, SUITE 720  
; CITY: PHILADELPHIA  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/353,700  
; FILING DATE: 09-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: REED, JANET E.  
; REGISTRATION NUMBER: 36,252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 563-4100  
; TELEFAX: (215) 563-4044  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HUMAN

US-08-353-700-1

Query Match 3.8%; Score 191.5; DB 1; Length 3248;  
Best Local Similarity 20.1%; Pred No. 0.00047;  
Matches 221; Conservative 179; Mismatches 400; Indels 301; Gaps 52;  
QY 6 KTVALTLTTSVVTHNQVFLSVKPEILKQTOASSISGADYAESGSKLKINETSQGV 65  
Db 2069 KTTALDQLSEKMEKKTQEL-----ESHQSECLHCIOVAEAVREKTELLQT---L 2115  
QY 66 DDTYTDLFSDKRTTPEKI-----KONLAKGPBEQELKAVTENTSEKQITSGSQLSQSKES 121  
Db 2116 SSDVSELLKDKTHLQELQSLQSEKOSQALSFKCELENOIAQNKBEKELL-----VKES 2168  
QY 122 LSLNKVTPSTSNWEICD-----FITKGNLTVLGLS-----KSGVEKLS---QTD 161  
Db 2169 ESLQARL-SESDYEKLVNSKALEAALVEKGFALRLSSTQBEVHQLRGIEKLRVRIAD 2227  
QY 162 -----HLV--LPSQAADGTQIQIVASFAPTPDKTKTAIAEYTSRAGENGIEISOLDVDG--- 211  
Db 2228 EKKQLHIAEKLKERENDSL-----KDKVENLERELQMSEENQELVILDAENSKA 2278  
QY 212 -----KEIINEGEVFNYSLLKKVTIPTGYKHI-----CQ-----DAF 243  
Db 2279 EVETLKTQIEEMARSLKIFE---LDLVTLRSEKENLTQIQEKGQQLSELDKLLSSFSL 2335  
QY 244 VDNKNIAEVNLPESLETISDYAFALHALKQIDLPDLNKAIGELAFDFNQITGKLSILPQOL 303  
Db 2336 LEEKEQAEIQKESKTAIVM-----LQNLKEL-----NEVAALCGDOBI 2377  
QY 304 MRLAEAFKSNHIKTIIEPRGNSLVIGEASQDNDLSQLMLPDGLKEKIESEAFTPNGDD 363  
Db 2378 MKATEQSLDPPIEBEHQLR-NSIEKL-RARLEADEKKQLCVLQQLKESHEHA-----D 2428  
QY 364 HNNRV-----VLWTKSGKNPSGLATENTYVNDPKSLWQSPEDITYTKWLEEDTYQKN 417  
Db 2429 LLKGRVENLERELRTARTNQEHAALAEANS-----KGEVETLKAKIEGTMQ--- 2474  
QY 418 SVTGFSGKGLQ-----KVKRNKNLEIPKQHNGVTITEIGDNAPRVDFQNKTLRKVDL 470  
Db 2475 -----SURGLELDVVVTRSEKENLTNELOKQERISELEIINSSPENT-LOEKQEKVQM 2528  
QY 471 BEVKLPSTIRKIGAFQFOSNNLKGFEASDDLEETK-EGAFMNNRLETLELQKLTIGDA 529  
Db 2529 KEKS--STAMEWLOTQLKELNERNVAALHNDQEAACKAEQNLSSQVECELEKALQLOGLD 2586  
QY 530 AFHINHIYAIVLPEVS-----QEIGRSAPFRONGANNLIFMGSKVKVTGLGEMAFLSNRLEHL 585  
Db 2587 EAKNYY---IVLQSSVKGLIQEV-----EDGKQKL-----EKKDEEISRLKNQIQ--- 2628  
QY 586 LSEQKOLTEIPVOAFSDNAL-KEVLLPASLKTIRREAFKQHLKQLEVASALSHTAFNAL 644  
Db 2629 --DQEQVSKLSQVEGEHQLWKEQNL--ELRNLTVLEQKIQVLQSKNAS-----L 2675  
QY 645 DNDGDEQPDNKKVVKTHNSVALADGHEFIVDPDKLSTIVDLKILKIEGLDYSLR 704  
Db 2676 QD-----TLEVLQSSYKNLENELELTMDKMSF-----VEKVNKM-----TAK 2713  
QY 705 QT-TQTPRDMTTAGKALLSKNSLRQGEKQKFLQEAQFPLGRVLDLKAIAKAELVTK 763  
Db 2714 ETELQREHMHM--AQKTAELQBEL-SGEKNRLAGELQLLLEI-----K 2754  
QY 764 ATKNGQLLERSINKAVLAYNNSAIKKA-----NVKRLKEKELDILTLGLVKGKGLAQATWVQ 819  
Db 2755 SSKD-QLKE-----LTLENSELKSLDCMHKQVKEGKGVREIEAEVQLRHEAEK 2806  
QY 820 GYVLLKTPPLPYEYIGLVNVDK-----SGKLIYALDMSDTIG 858  
Db 2807 QALLLDTN---KOYEVEIQTREKLTSECECLSSQKLEIDLKSSKEELNNSLKATQTIL 2863  
QY 859 EQOKDA-----YGNPILNVDDNNGYHALAVATLADYEGLDIKTILNSKLSOL----- 906  
Db 2864 BELKTKMDNLKYVNLKKNENRAGQKMKLLKSCQLE--BEKEILOKELSOLOAAQEK 2921



```
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqFormat Version 0.9
; SEQ ID NO 1154
; LENGTH: 3210
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P49454
US-09-538-092-1154

Query Match      3.7%; Score 187.5; DB 4; Length 3210;
Best Local Similarity 20.1%; Pred. No. 0.00089;
Matches 221; Conservative 179; Mismatches 400; Indels 301; Gaps 52;

QY      6 KTVALTLTVTSVVTHNQEVPSLVKPEILKQTQASSSISGADYAESSGKSKLKINETS GPV 65
DB      2069 KTTALDQSEKMKETQEL-----ESHQSECLHICIQVAEAEVKEKTELLQT---L 2115

QY      66 DDTVTDLFSDXRTTPEKI---KDNLAKGPREQELKAVTENTSEKQITSGSQLEQSKES 121
DB      2116 SSDVSELLKDXHLQKLEKQSLQKLSQALSLTKCELENQALNKEKEL-----VKES 2168

QY      122 LSLINKTVPSTSNWEICD-----PITKGNLTVLGLS-----KSGVEKLS---QTD 161
DB      2169 ESIQARL--SESDYEKLNVSKALEAALVEKGEFALRLSSQTQEEVHQLRRGIEKLVRREAD 2227

QY      162 -----HLV--LPSQADGTQLQVASFAPTPDKKTAIAYTSRAGENGESIQLDVDG--- 211
DB      2228 EKKQLHIAEKLKERENDSL-----KQVLENLRELQWSEENQELVILDAENSKA 2278

QY      212 -----KEIINEGEVNSVLLKKVTIPTGYKHI-----GQ-----DAP 243
DB      2279 EVETLTQTOEEMARSLKVEF---LDLVTLRSEKENITKQIQKQQLSELKLLSFKSL 2335

QY      244 VNKNIAEVLNPESLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSPLQRL 303
DB      2336 LSEKEQAEIQIKESKATAVEM-----LQNLKEL-----NEAVALCGDQEI 2377

QY      304 MRLAERAPKSNHIKTIEFGNSLKVIGEASFQNDLSQLMPLDGLKIESEAFNGPDG 363
DB      2378 MKATEQSLLDPPIEEHQLR--NSIEKL--RARLEADEKKQLCVLQQLKESHH-----D 2428

QY      364 HYNRRV-----VLWTKSGKNPGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKN 417
DB      2429 LLKGRVENLERELEIARTQEHAALEAENS-----KGEVTLKAKLEGTQ--- 2474

QY      418 SVTGF SNKGLQ-----KVRKNKNLEIPKQHNGVTTITGDNAPRNVDFONKTLRKYDL 470
DB      2475 -----SLRGLEDVVITRSEKEDLTNQLQEKERISELEI INSSPENI--LQSEKEQKVQM 2528

QY      471 EEVKLPSTTRKIGAFAPQSNVLKSPFASDDLEIK--EGAFPMNRITETLEKOKLVITGDA 529
DB      2529 KEKS--STAMEMLQTLQELNERNVAAHNDQEAACKAQNLSSQVCELCLEBAQLLQGLD 2586

QY      530 AFPHINHYAIVLPESV---QELGRSAPFGNGANNLI FMGSKVKTLGEMAFISNRLHLD 585
DB      2587 EAKNNY---IVLOSSVNGLIQEV-----EDGQKL-----EKDDEISRLKNQIQ--- 2628

QY      586 LSEQKQLTEIPVQAFSDNAL--REVLIPASLKITIREEAFKKNHLKQLEVASALSIAFNAL 644
DB      2629 --DQEQVLVKLSQVEGEHLWKQENL--ELRNLTVELEQKIQVLSQKNAS-----L 2675

QY      645 DNDGDGEQDFNVKVVVTHNSALADGEHPIVDDPKLSSTIVDLEKILKIEGLDYSTUR 704
DB      2676 QD-----TLEVLQSSYKNLENELELTAKMDKMSF-----VEKVNQK-----TAK 2713
```

```
QY      705 QT-TQTQFRDMTTAGKALLSKNSLRQGEKQKFLQBAQFPFLGRVLDLKAIAKAEKALVTYK 763
DB      2714 ETELQREMHM--AQKTAELQBEL--SGEKNRLAGELQLLLEI-----K 2754

QY      764 ATYNGOLLERSINKAVLAYNNSAIKKA-----NVKLEKELDLTLTGLVBEKGPPLAQATMVQ 819
DB      2755 SSKD-QLKE-----LTLENSELKSLDCMKHQVKEKGVREIARYQLRLHEAEKKH 2806

QY      820 GVVLLKTLPLPEYIYGLNVVFDK-----SGKLIYALDMSDTIG 858
DB      2807 QALLLDTN---KQYVEIQTREKLTSGECLUSSOKLEIDLKSKSEELNLSLKATQIL 2863

QY      859 EQQKDA-----YGNPILNVDEDNHYHALAVATLADYEGLDIKTILNSKLSQL----- 906
DB      2864 BELKTKMDNLKYNVQLKKENERAQGMKLLIKSKCKQLE--BEKBITQKELSQLQAAQEK 2921

QY      907 -----TSIRQVPTAAYHRAGIFQATQNAABAEQLLPKG-----THSEKSSSESAN 954
DB      2922 QKTGTVMDTKVDELTTTEKE---LKETLEEKTKEADEVLDKYCSLLISHKLEKAKEMLE 2978

QY      955 SKDRGLQSNPKTNRGRHSAIL 975
DB      2979 TOVAHLCSQOQSKQDSRGRSPLL 2999

RESULT 17
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match      3.6%; Score 185; DB 4; Length 3878;
Best Local Similarity 19.5%; Pred. No. 0.0018;
Matches 180; Conservative 164; Mismatches 349; Indels 230; Gaps 44;

QY      7 TVALTLTTVSVVTHNQEVPSLVKPEILKQT--QASSSISGADYAESSGKSKLKINETS GP 64
DB      94 TIMRTLHSGEITSHEQG-FSVELESEISTTADDCSSEVNGCSFVVRTGKPTNLLREERG 152

QY      65 VDDTVTDL-FSDKRTTPEKIDNKLAKPREQELKAVTENTSEKQITSGSQ-LEOSKESL 122
DB      153 VDDSYEQGAQDSPTHELMWSELA--GKQHEIEELNRELE-EMRVTYTEGLQQQLQEF 209

QY      123 SLNK-----TVPSTANWEICDPITKGNLTVLGLSKSGVEKLSQTDHLVLPDQADGTQLIQ 177
DB      210 AAIKQDGGIITOLTANQOAR--REKDETM---REFLELTEQSKQLQIQFOQLQASSETLR 264

QY      178 VASFAFTP-----DKTATAEYTSRAGENGESIQLDVDGKEIINEGEVFN---SYLLKKVY 230
DB      265 NSTHSTAADLLQAKQOQLTHQOQLEEQDHLE-DYQKKK-----EDFTMQISFLQEKIK 318

QY      231 IPTGYKHICQDAFVNKNIAEVLNPESLETISDYAFALHAKQIDLPDLN-----KAIGEL 286
DB      319 V-----YEMEQDKVENSKEBIEQEKETIEELNTKIIIEEKKTELKOKLTADKLGL 374

QY      287 AFFDNOITG-----KLSLP-----RQLMRLAERAFKSNHIKTIEFGNS 325
DB      375 ---QBQIVQKNQIBIKNMKLELTNSKQKERSSEIKQLMGMTVEELQKRNH-KDSQF----- 426
```













```
Db 208 NTGALTDTSQYQFKQ-----GLTKLNDYTPHNQIVNFEN--TSLETIDNYVTADSWYRP 260
Qy 269 -----LALKQIDL-----PDNLKAIGELAFPDNQITGKLSLPRQLMLAEBA 310
Db 261 KDILKNGKWTWASESDLRPLMSWMPDKQTOIAYLNYMNQOGLG-----TGENY 310
Qy 311 FKSNIHKTIIEFRGNSLKVGEASFQNDLSQLM-----LPGLEKIESEAFPTGNP 360
Db 311 TADSSQESLNLAQTQVQKIEKISQTOQTQWLRLDIINSFVKTPQPNWSQTESDTSAGE- 369
Qy 361 GDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESP--EIDYTKWLEBDFTYQKNS 418
Db 370 -KDHLOGGALLYSNSDK-----KNAINHLSIL-----EAWSDNDPOY---NKDTKGAQL 413
Qy 419 VTGFSNKGLOKVKRNKNLEIPKQHNQVITTEIGDNAPRVNDFONKTLRKDYDLEEVKLPST 478
Db 414 -----SSGGYDFLLAN-----DIDNSNPVVOAEQLN-----439
Qy 479 IRKIGAFQSNLKSFEASD-----DLEEIKEGAFMNNRIETLEKDKLVTIGDAAFHINH 535
Db 440 -----WLHYLMNYGSIIVANDPEANFDGVRVDVAVDNNVADLLQI-----ASDYLKA 484
Qy 536 IYAIVLPSVQESIGRSAFRQNGANNLIIPGSKVKTLGEMAFLSNRLEHLDLSEQKLTETI 595
Db 485 HYGVDKSE-----KNAINHLSIL-----EAWSDNDPOY---NKDTKGAQL 521
Qy 596 PVQAFSDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSASHIAPNALDDNDGD-----E 651
Db 522 PI-----DNKRLSLLYALTRPLEKDSNKN-----EIRSGLEPVITNSLNNRSAGEKNSE 572
Qy 652 QFDNKVVVKTTHNSY--ALADGSHFIVDP--DKLSSTIVDLEKILKLEGLDYSTLRQTT 707
Db 573 RMANYIFIRAHDSFVQTVIAKIIKAQINPKTDGLTFIDELKQAFKIYNEWMROAKKYYT 632
Qy 708 QTOFRDMTTAGKALLSK-----SNLRQGEKQKFLQEAQFGLGRVLDLKAIAKAFKALVTYK 763
Db 633 QS-----NIPTAYALMLSNKDSITRLYYGD--MYSDDQYMATKSPYYDAIDTLKKARI--K 685
Qy 764 ATKNGOLLERSI---NKAVLAYNNSAIKKNVKNLEKELDLTLGLVEGKPLAQATWVOG 820
Db 686 YAAGQDMKITVYEGDKSHMDYTGVLTS-----VRYGTGANEATDQGEATKTQG 737
Qy 821 VYLLKTPPLPEYIIGLVVYFDKSGKLIYALDMSDTIGEGOKDAYGNPILNVDENEGYH 880
Db 738 MAVITSNPN-----SLKLNQNDKIV-----NWGAHKKQYERPLILLTTKGLTSY 783
Qy 881 ALAVATLADY-----EGLDIKTILNSKLSQTSIRQVPTAAYHRAGIFQAIQNA 929
Db 784 TSAPAAKSLYRKTNDRKGLVFDASDIQGYLNPQVSGYLAV-WVPVCGASNDQVVRVAASNK 842
Qy 930 ABAEAOQLPKPCTHSEKSSSES-----ANSKDRGLQSNPKTNR 968
Db 843 ANATQVY-----ESSALDSQLIYEGFSNFQDFVTKDSYTNK 881
```

## RESULT 25

```
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740, 274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
```

```
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6
```

Query Match 3.5%; Score 176; DB 4; Length 1430;

Best Local Similarity 17.9%; Pred. No. 0.0017;

Matches 191; Conservative 180; Mismatches 401; Indels 294; Gaps 43;

```
Qy 1 MKKHLKTLVALTLTTVSVVTHNQEVFSLVKEPILKQTOASSISGADYAESGSKLKINE 60
Db 12 VKKHVVTVAVASGLITLGT-----TTLGSSVSAETEQTQS-----46
Qy 61 TSGPVDVTVDLFSKRTTPEKIKNLAKGPRBQELKAVTENESKQI--TSGSOLEBSK 119
Db 47 -----DKVVTQKSEDDKAASESSQTDAPKTKQAQ-----TEQTAQSOANVADTSTITK 96
Qy 120 ELSLANKTVPTSTSNWEICDFITKGNLTVGLSKSGVEKLSQTDHLVLPQAAAGTQLIQVA 179
Db 97 ETPSNITTOQANS-----DDKTVTNTKSEEAQTSEERTKQSEE-----AQTTASSQAALTOA 147
Qy 180 SFAFPTDKKTAIAB-----YTSRAGENGESISOLDVDGKEII---215
Db 148 KAECLKQRTAAQENKNPVDLAAIPNVQIDGKYIYIGSDGQPKKNFALTNNKVLVYFPK 207
Qy 216 NEGEVENS--YLLKKVTTIPTGYKHIGODAFVDNKNIAEVLNPESLETTISDYAPAH---268
Db 208 NTGALTDTSQYQFKQ-----GLTKLNDYTPHNQIVNFEN--TSLETIDNYVTADSWYRP 260
Qy 269 -----LALKQIDL-----PDNLKAIGELAFPDNQITGKLSLPRQLMLAEBA 310
Db 261 KDILKNGKWTWASESDLRPLMSWMPDKQTOIAYLNYMNQOGLG-----TGENY 310
Qy 311 FKSNIHKTIIEFRGNSLKVGEASFQNDLSQLM-----LPGLEKIESEAFPTGNP 360
Db 311 TADSSQESLNLAQTQVQKIEKISQTOQTQWLRLDIINSFVKTPQPNWSQTESDTSAGE- 369
Qy 361 GDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESP--EIDYTKWLEBDFTYQKNS 418
Db 370 -KDHLOGGALLYSNSDK-----KNAINHLSIL-----EAWSDNDPOY---NKDTKGAQL 413
Qy 419 VTGFSNKGLOKVKRNKNLEIPKQHNQVITTEIGDNAPRVNDFONKTLRKDYDLEEVKLPST 478
Db 414 -----SSGGYDFLLAN-----DIDNSNPVVOAEQLN-----439
Qy 479 IRKIGAFQSNLKSFEASD-----DLEEIKEGAFMNNRIETLEKDKLVTIGDAAFHINH 535
Db 440 -----WLHYLMNYGSIIVANDPEANFDGVRVDVAVDNNVADLLQI-----ASDYLKA 484
Qy 536 IYAIVLPSVQESIGRSAFRQNGANNLIIPGSKVKTLGEMAFLSNRLEHLDLSEQKLTETI 595
Db 485 HYGVDKSE-----KNAINHLSIL-----EAWSDNDPOY---NKDTKGAQL 521
Qy 596 PVQAFSDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSASHIAPNALDDNDGD-----E 651
Db 522 PI-----DNKRLSLLYALTRPLEKDSNKN-----EIRSGLEPVITNSLNNRSAGEKNSE 572
Qy 652 QFDNKVVVKTTHNSY--ALADGSHFIVDP--DKLSSTIVDLEKILKLEGLDYSTLRQTT 707
Db 573 RMANYIFIRAHDSFVQTVIAKIIKAQINPKTDGLTFIDELKQAFKIYNEWMROAKKYYT 632
Qy 708 QTOFRDMTTAGKALLSK-----SNLRQGEKQKFLQEAQFGLGRVLDLKAIAKAFKALVTYK 763
```

Db 633 QS---NIPYALMLSNKDSITRLYYGD---MYSDDQGYMATKSPYYDAIDTLTKARI--K 685  
Qy 764 ATKQGLLERSI---NKAVLANNNAIKKANYRLEKELDLTLGLVEKGPLAQATWVG 820  
Db 686 YAAAGQDMKITTYVEGDKSHMDWYTGVLTS-----VRYGTGANEARDDQSEATKQG 737  
Qy 821 VYLLATPLPLPYEYGLNVPYDKSLIYALDMSDTIGEGQKDAYGNPTLNVDEDEGCVH 880  
Db 738 MAVITSNNP-----SUKLQNDKVIV-----NMGAHNKQYRPLLLTTKDLGTSY 783  
Qy 881 ALAVATLADY-----EGLDIKTLNLSKLSQTSIROVPTAAVHRAGIFOAIONA 929  
Db 784 TSDAAKSLYRKTNDKGLVFPDASDIQGYLNPQVSGYLAV-WVPVGSNDQDVRVAASNK 842  
Qy 930 AAASQLPLKPTHESEKSSSSS-----ANSKDRGLQSNPKTNR 968  
Db 843 ANATQVY-----ESSSALDSQLIYBGFNSFQDFTVKDSYTNK 881

RESULT 26  
US-09-248-796A-20284  
; Sequence 20284, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 20284  
; LENGTH: 1078  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-20284

Query Match 3.5%; Score 175.5; DB 4; Length 1078;  
Best Local Similarity 20.9%; Pred. No. 0.0012;  
Matches 185; Conservative 142; Mismatches 326; Indels 231; Gaps 42;

Qy 9 ALTLTVSVVTHNRVF-----SLVKEPILKQTOASSISGADYAESGSKSLKIN-ET 61  
Db 308 ALTLTFSQDATGLPTVYLTVDVYVNLVKDNFLR-----IKRALFHPDPSBPRGKISYEL 360  
Qy 62 SQPVDDTVTDLPSDKRTTPEKIKDN-----LAKGPREOELKAVTEN 102  
Db 361 FEELDTKVNLTKELQTEKENAESNDKELNEKIEKLTNLSTKLETKLEDKEQELAKIQED 420  
Qy 103 TS--SEKQITSSQLEQSKESLSLNKTVFSTSNWEICDFITKGNLTVGLSKSGVKLSQT 160  
Db 421 HKLLNEKFLVTTNSLSAMKASKEFETASQYKQKLEQALKXGNTSESTLKQLEKLNST 480  
Qy 161 D-----HL-VLPQADGTQLIQ-----VASFAFTPDKKTAAIABYT 195  
Db 481 EQAKKLEDGINMTDRDLFLHKLKASEAE--TQIKQREREFKGLTVLENTKKDYELQIN 538  
Qy 196 SRAGENGEISQ-----LDVDGK-----EINEGEVFNYSLLKKVITPTGYKH 237  
Db 539 NLNKNNEFPKQINELSKIESLTEDNKSNAKQLEKRLDTEENHEHMLDKLRAS----- 594  
Qy 238 IQGDAFVDMKNIAE---VNLPSLETISDYAFAPHALAQIDLPDNL-KAIGELAFDNOI 293  
Db 595 VAYNDLKXAKSSEETVAKSELETLSKI-----DNLEKELKEQQSKNEL 642  
Qy 294 TKLSLPLQMLAERAFK--SNHIKTIEFRGNLSKVIGEASFQDNLSQLMPLDGLKXI 351  
Db 643 EQGL---QNIITDSTNEKFKLEDELKSIK-----KSNKEISSQNSLIQ-----KLEKT 688

Qy 352 ESEFTGNPGDDHYNRVRVLWTKSGKNPGLATENTYVNPDKSLWQSPSEIDYTKWLEED 411  
Db 689 EKOL-----QAKDEIDKLKAEK-----SNIDNUNSEISSQSKLKEA-----BES 730  
Qy 412 FTYQKNSVTGFSNKGLOKVK-----RNKNLETPKHQHGVTIPEIGDNAPRNVDFONKTLRK 467  
Db 731 HSSTKDEHSSLS-ENLKKLKEEYENTKTSMLAK-----LSAKIBEHKKATDEIETKTKHI 784  
Qy 468 YDLBEVKLPSTYRKGIFAPFOSNNLKS-----PEASDDLEBI-----KEGAFMNNRIETL 517  
Db 785 TDLOE-----EHAQKQSFESERNDIKSNLDEANKELSDNREKLSNLEKEKTELANKLKTQ 840  
Qy 518 ELKDKLVTIGDAAFHINHIYAIVLPESVQETIGRAFQNGANNLIFMGSVKVTILGEMAF 577  
Db 841 BEK---ISDLETSVAISEDKSKUKHIEDLKR-----EKIKL---ETTLK 880  
Qy 578 SNRLEHLDLSHQKQLTEIPVQAFSDNALKEVLLPASKLTIRE-----EAPKK 624  
Db 881 ENBESTMFEKKEQLQVV-----NDKCKE--LEACLKLTETKEKEINDLIRKLEAKS 930  
Qy 625 NHLKQLEVASALSIAFNALDNDGDQFQDNKVVVTKTHNSYALADGSH----- 673  
Db 931 DHDTERKKLSLL-----IEDTKSEB---KNVILKNEQIEKL-KGSEKEVRDIQSQL 979  
Qy 674 --FIVDPDKLSTIVLEKILKLEGLDYSLTROTTOFQDMTAGKALLSKSNLRQGE 731  
Db 980 AAKTTDWEKIKTT---LDKVLKESDLE-KTNKESVDTLKKEVENLKKEISLLEDQKKDD 1035  
Qy 732 KQKFLQ-EAQPFLGRVOLDKAIKAEKALVTKKATNGQLLERS 774  
Db 1036 TTKYKELAAQLSTKTSNLDSTTMELEKTELELKKVRN-ELTEVS 1078

RESULT 27  
US-09-583-110-3824  
; Sequence 3824, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; LENGTH: 1216  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-3824

Query Match 3.5%; Score 175.5; DB 4; Length 1216;  
Best Local Similarity 19.4%; Pred. No. 0.0015;  
Matches 208; Conservative 133; Mismatches 380; Indels 349; Gaps 46;

Qy 96 LKAVENTESKQITSSQLEQSKESLSLNKTVFSTSNWEICDFITKG-NLTVGLSKSGV 154  
Db 1 MKLIPFLSEEBIQLOEAEANSKEQ---KKTAE-----QIEAITYTSQNTLVASASGSG 52  
Qy 155 EKLQTDHLVLPQAAQDTQLIQVASFAFTPDKKTAAIABYTSRAGENGEISQLDVDGKEI 214  
Db 53 KTFWVERIL--DQARGVEISQLFISTFTVKAATELKE----- 89  
Qy 215 INEGEVFNYSLLKKVITPTGYKHIGQDAFVDMKNIAEVLNPSLETISDYAPAHALKQI 274  
Db 90 -----RLEKKIS-----KKIQETDDVDLQK-----HLGRQLA 116





/ APPLICANT: Sakowicz, Roman  
/ APPLICANT: Goldstein, Lawrence S.B.  
/ APPLICANT: Cleveland, Don W.  
/ APPLICANT: The Regents of the University of California  
/ TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for  
/ TITLE OF INVENTION: Chromosome Congression  
/ FILE REFERENCE: 18557C-000110US  
/ CURRENT APPLICATION NUMBER: US/09/150,867  
/ CURRENT FILING DATE: 1998-09-10  
/ EARLIER APPLICATION NUMBER: US 60/058,645  
/ EARLIER FILING DATE: 1997-09-11  
/ NUMBER OF SEQ ID NOS: 11  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 1  
/ LENGTH: 2954  
/ TYPE: PRT  
/ ORGANISM: Xenopus sp.  
/ FEATURE:  
/ OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)  
/ OTHER INFORMATION: member of the kinesin superfamily of microtubule  
/ OTHER INFORMATION: motor proteins  
/ FEATURE:  
/ NAME/KEY: DOMAIN  
/ LOCATION: (1)..(472)  
/ OTHER INFORMATION: kinesin like motor domain  
/ FEATURE:  
/ NAME/KEY: DOMAIN  
/ LOCATION: (473)..(2752)  
/ OTHER INFORMATION: rod domain  
/ FEATURE:  
/ NAME/KEY: DOMAIN  
/ LOCATION: (2753)..(2954)  
/ OTHER INFORMATION: tail domain  
/ US-09-150-867-1

Query Match 3.4%; Score 173.5; DB 4; Length 2954;  
Best Local Similarity 20.1%; Pred. No. 0.0076;  
Matches 209; Conservative 186; Mismatches 431; Indels 213; Gaps 47;

QY 23 EYFSLVKEPILKQTOASSISGADYAEBSGSKSLKINETSQVDDTVTLDFSKRTTPEK 82  
DB EVKLLLEMLLGHLLTDSLS---IEKLQLENLEVTETKLTQLEBEMKNITIERNELQTN 1255  
QY 83 IKDNLAKGPREGELKAVNTSESEKQITSGSQLEKSKLSLNTKVPSTSNWEICDFTYK 142  
DB 1256 FED--LKAHDSLKQDLSNI--EQSIETQDSLRAAQBELREQQLVDSFRQQLLD--- 1307  
QY 143 GNTLVGLSK-----SGVEKLSQTHLVLPQAAQDQO-LIQVASFAPTPDKKTAIEVT 195  
DB 1308 --CSVGISPNHDVANQKVKLGEVNSLQSEMLRGERDELQTSKALVSELELRAHVK 1365  
QY 196 SPAENGEIS-QLVDGKEIINEGEVFNYSLLKVTIPTGYKHIGQDAFVDMKNIAEYN- 253  
DB 1366 SVEGENLEITKGLNGLEIKLSE--ESEVLKSM-----ENLKEDNN 1407  
QY 254 -LPSLETTSD-----YAPAHALKQID-LPDMILKA-----TGLAFPD----- 290  
DB 1408 KLKQAEYSYSSKENQFSLEEVFSGSKLVDEIVLKAQKAAERLEIKORDYFELVQTA 1467  
QY 291 --NQITGKLSLRLMLRAERAPKSNHIK-TTEFRGN--SLKVIAGEASFQDNDL----- 339  
DB 1468 NTNVLVEGKLETPLQ-----ADHEEDSIDRSEEMIKVLGEKLERNGYLLERLQE 1517  
QY 340 SOLMLPDGLEKTESAFTGNPDGHYNNRVLTWTKSGKNPSGLATENTYV--NPDKSLMQ 397  
DB 1518 EKLESLNKLLEIKQEMETS-----VLLKDDLLQOKLESLSLSENIILKENIDITLKH 1567  
QY 398 ESPEIDYTKWLEDEFTYQNSVTGSGNKG--LQVKRNKNLEIPQHNGVTITEGDNAF 455  
DB 1568 HSDTQAQLOKTOELQAKNLAAASDNCPIQTEKETSADCVHPLEEKILLTLE---ELH 1624  
QY 456 RNVDLFQNTL-RKYDLE--EVKLPSPIRKIGAPAFQS--NNLKSFE--ASDDELEIKEGAFM 510

DB 1625 QKTNEQEKLLHKEKNELEQAQVELKCEVHELMKSMIESKSSLESLOHEKCHDTEQQLLALKQ 1684  
QY 511 NNRIETLEKOKLVITIGDAAPHINHIYAIULPESVQEIIGRSAPFRONGANNIIFMGSKVKT 570  
DB 1685 QMQVVTOEKKELOQTHLTAEDVDH-----LKENI-ELGLN-FKNE-----AQOKT 1728  
QY 571 LGEMAFLSNRLEHLDLSEKQQLTRIP--VQAFSDNALKEVLLPASLTIREEAPKKNHLK 628  
DB 1729 TKEQCLL--NENKELEQSOHRLQCEIEELMKSLKD---KE-----SALETKESEOKVINLN 1780  
QY 629 Q-----LEVASALSHIAFNALDNDGDQDFONKVVVTKTHN-----SYALADGEHFIV 676  
DB 1781 QEMEMVMEELKNSORTVIAERDQLQDDLRRESVENSIEQTODDLRKAQALQOKKQKV- 1839  
QY 677 DPDKLSSTIVDLKILKIEGLDYSLRQTQTOTFRDMTTAGKALLSKNSLRQEKQKXF- 735  
DB 1840 --QELTSQISVLQEKISLLE-----NOMLYNVATVKETLSERDDLNSQSKHLFS 1886  
QY 736 -----LQEAQFPLGRVLDLKAIAKAEKALVTKKATK-NGOLLERSINKAVLAYNNS- 785  
DB 1897 BIETLSLSLKEKEFALEQAEKADAAKRTIDITEKISNIEBEQLLQQAATNKETIYERES 1946  
QY 786 -----AIKKANVKRLEKELDLLTGLVEGKGLAQAATMVQGVYLLKTPPLPEYIIGL 837  
DB 1947 LIQKEQLALNTEHLRETLKSKOLALGKMEQERDEA---ANKVIALTEKMSLEEQINE 2002  
QY 838 NVYFDKSG-----KLIIY-----ALDMSDTIGEQQDAYGNPILNVDR-----DNEGVA 881  
DB 2003 NVTTLKEGEKETFYLRPSKQOSSQMBELRESLTKTDIQLBEAEKEISEATNEIKNL 2062  
QY 882 LAVATLADYEGLDIKTLNLSKLSQTSIRQVPTAAYHRAGIPQAIONAAAEBQLPKPG 941  
DB 2063 TAKISSLEEELQNASILNEAVSERENLR-----HSKQQLVLSLEQL--SLT 2107  
QY 942 THSEKSSSESANSKDRGL 960  
DB 2108 LKSRDHAPAKREKDEAV 2126

RESULT 30  
US-09-248-796A-16333  
; Sequence 16333, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 16333  
; LENGTH: 905  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-16333

Query Match 3.4%; Score 172; DB 4; Length 905;  
Best Local Similarity 19.9%; Pred. No. 0.0017;  
Matches 178; Conservative 148; Mismatches 328; Indels 242; Gaps 42;

QY 19 THNQEVSFLVKE-----PILKQTOASSISGADYAEBSGSKSLKINETSQVDDTV 69  
DB 41 TNSLSNQLLSAKVLATETPDVSKLTQPSDEEISNNVSTEDGSEK-----ATLPPINNT 96  
QY 70 TDLFSDKRKT-----PEKIKDNLAKGPRBOELKAVNTESKQITSGGOLROSK 119  
DB 97 TDTTDTNGTATTTAAAAAQAQSDLDQL-PAPIKSLKKFAKBEKYPILLDAVYKIEKK 155



```
Db 1503 QLEKAKKEGAVINVAALSAG--TDKSGAGVAIAVNTVKNFKPAELSGSNKEAGEDKIH 1560
Qy 801 LITGLVEGKPLAQATMVGQVYLLKTPLPPEYIYGLNVYFDKSGKLIYALDMSPTIGEG 860
Db 1561 AKHVNVAKSSVTVVNAASGLAISKDAFS---GMSGAWQDLS-----NDTIAKV 1607
Qy 861 QKDAVGNPILNVDENE-----GYHALAVATLADYEGLDIKIL----- 899
Db 1608 DKGRISADSLNVNANSILGVNVVAGTIAAGSLSTAVGAAPANTLNKTSALITGTVKNPF 1667
Qy 900 ---NSKL-----SOLTSIRQVPTAAVHRAGI--FOAIONAAAEOLLPPKPGTHSB-K 946
Db 1668 SKNTKVNQALNDSHITNVSAGGASIKQAGIGWVSNRGSDEALV-----SDSEFE 1723
Qy 947 SSSSESANSKORGLSNPKNTRGRHSAILPRTGSKGSFYVIGLTVTSVALLSLITAIKKK 1006
Db 1724 GVSSFNVDAKO---QKTINTIAGN-----ANGGKAAAGVATVAHTNIGKQSVIAIVKNS 1774
Qy 1007 K 1007
Db 1775 K 1775

RESULT 32
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380170
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5178

Query Match 3.3%; Score 169.5; DB 3; Length 1010;
Best Local Similarity 20.5%; Pred. No. 0.003;
Matches 164, Conservative 116; Mismatches 284; Indels 235; Gaps 38;
Qy 19 THNQEVFLVKBPIKQTOQASSIS-----GADYAESGSKSKLKINETSGPVDD--- 67
Db 215 TPNNDALALYKE--LESSQTDKMIKFPQFNDYGCKILKSPKAKNKITK---ELDDLNH 269
Qy 68 ----TWTDLFSDKRTTPEKIK-DNLAKGPR-----EOLKAVTENTSEKQITSSQLEQ 117
Db 270 KYKVNVELSENTKGLKAEIKFDDLKKEQYIDKLQELKMIQESKVLITYFTRLQSLKK 329
Qy 118 SKESL-----SLNKTVPSTGNWEICDPTTKGNTLVGLSKSGVKLSQTDHLVLPQAA 170
Db 330 DKDELVSLEHQSKLNET-----NYH-----NEIKGFQK-----QLEHL----- 362
Qy 171 DGTQLIQVASP-----AFTPDKKTAIAYTSRAGENGESQLDVGKGEIINSEGVFN 222
Db 363 -STRENEITQFNQLEKQNVFNQDLKIISSYQKQPVIEEIKRLYSEVNDLITKKEBLT 421
Qy 223 SYLLKKVITPTGYKIGQDAFVDNKNIAEVNLPESLETISYAPFAHLAKQIDLDPNLKA 282
Db 422 KEMNNKNDFAITIEHYTEIYKXKIIIDESERQKDEKLFDKL-----QLDKSSYLSK 474
Qy 283 IGEAFFNDQITGKL-SLPLQMLRAERAFKFNHKK-----TIEFRGNSLKVIGE--- 331
Db 475 LKEKKEQLNEIBESSITNIDATLIDNDKQDFVNEIKSAMSIGDTFCICGNEIHSIGEHID 534
```

```
Qy 332 ---ASFQNDLSQL-----MLPDGLEKIESAFTGNPGDDHYNRVRVLWTYSGKNPQSLA 383
Db 535 FESIAQNNKIKRLSKKVKIRDBIIKTIETRIEELN-----HRENELAFKOEKDKDISBLQ 590
Qy 384 TENVYVNPDKSLMOESPEIDYTKWLEBDFTYQKNSVTG-----FSNKGLOKVKRKNK 435
Db 591 KOLNHLN---OLKDEQSSINK---LVENFEKQEKEIVNKHQFDLDSRRKWTQKEK----- 640
Qy 436 LEIPKQHNGVITBIGD---NAFRNV-DFQ-----NKLTKYDLSSEVKLPSTIRKIGA 484
Db 641 LEI-----QINDFERHSQFSSVNDFTYYSHAKQVETVEYENEKTKDLNEL--- 688
Qy 485 FAFOSNNLKSFEASDDLEEEIKGAPMNRITLBLEKOKLVLTIGDAAAPHINHIYALVLPES 544
Db 689 ---NNKLKTEMNDQKHLTENLTQTSKEINNLEK-----MEKE 723
Qy 545 VQETGRSAFRONGANNLIFMGSKVKTLGEMAFLSNRLEHLDLSEQKLTETIPVQAFSD--- 602
Db 724 MQLGFESYDO-----VKSAA-----DLSAQKDBEIEREINIYNKY 759
Qy 603 -----NALKEVLLPASLKTIRE--BAFKKNHLKQLEVASALSASHIAFNALDDNDGD--- 650
Db 760 QSYEIEINRLKELVKGKLLNLEBLRQSIETKLNKLDETNSQIATISYK-IDNNSNKNK 818
Qy 651 -----EOPDNKVVVKTTH-----HNSYALADGEHFIVDDPKLSTIVDLLEKI----- 691
Db 819 IKNIQIOLDLDELKVKQKEIFLLSEILAGKNDYKLT-LENYVL-----IYYLEKLIIPQA 869
Qy 692 ---LKLLEGLDYSTLRQTT 707
Db 870 NQRLSFMGNGRYQIRRET 888

RESULT 33
US-09-583-110-3829
; Sequence 3829, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3829
; LENGTH: 1965
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-3829

Query Match 3.3%; Score 169.5; DB 4; Length 1965;
Best Local Similarity 19.4%; Pred. No. 0.0079;
Matches 231, Conservative 167; Mismatches 442; Indels 351; Gaps 56;
Qy 4 HLKTVALTTLTVSVVTH-----NOEVFSLVKPEILKQTOQASSISGADYA--- 48
Db 136 HLVLVAGIGNGVLSVVALENHLLNVTVDYELTSGRKLPLPKE-----ISGVTYIGYI 189
Qy 49 -----ESSGSKLKINET-SGPVDDTVDLFSDKRTTPEKIKD-----N 86
Db 190 KEGKTTSDFEVSNQEKSAATPTKQKVDYNTVFNFDHPSTVQAIQEQTPVSTKPTBEVQ 249
Qy 87 LAKGPREQEL-----KAVTENTSEKQITSGQSKESLSLNKTVPSSTNWEICDFTTK 142
Db 250 VVEKFPSTELINPRKEEQSSDSQELAEHKNLETKKE-----EKISPK----- 293
```

```
QY 143 GNTLVGLSKSGVEKLSQTDHLVLPQQAADGTOLIQVAFPTDPKKTAIAEYTS----- 196
Db 294 -----EKTGVNTLNPQDE-VLSGQLNKPPELLYREETIEIKIDFQEEIQENPDIVGIV 345
QY 197 RAGENGESQLDVGKEI-----INGEVFNYSLLKKVTIP----- 232
Db 346 RVKQEGKL-----GKKVEIRVIFSVNKEBVSREIVSTSTTAPRIVEKGTGKTQVIKE 399
QY 233 ---TGKHHIG-----QDAFVONKNIAEVLNLPESLETISDYAFALHAKOI 274
Db 400 QPETGVEHKDVGSGAIVEPAIQPELFEAVVSDGKGVPEVPQALSKAVITDKGETEV---QP 456
QY 275 DLPDLNLKAIGELAFFDNQITGKLSLPRQLMRLAERAFKSNHIKTIIFRGNLSKLKVGESAF 334
Db 457 ESPDT-----VVSDDGPEQVAPLP-----EYKGNIEQVPEPTV 491
QY 335 QNDLSQLMLPDGLEKIE-----SEAFGTNGPDGHHNNRVVLWTKSGKNPGLATENT- 387
Db 492 EKT-----EQPEKTEVPVKPTBETVPNNEGTEGTSI---QEAENPVQABESTT 542
QY 388 ---YVNPDKS---LWQES-----PEIDYTKWLEEDFTYQKNSVTGFSNKG 426
Db 543 NBEKSPDTSSENTGESSNPDSTTSVGESKNPEHNDKNSKENSEKTVEEVPVN--PNEG 600
QY 427 LQKVKNKNLEIPKQHNQVITITEIGNAFRNV-DFONK-TLRKYDLEEVKLPSTIRKIGA 484
Db 601 TVEGTSNQETEPQVPAETQNTSGKIANENTGEVSNKPSDKPPVEESNQ-----EKNGT 657
QY 485 FAFQSNLKSFEASDDBLEIKGAFNWNRIETLEKDKLVITGDAAFINHIYAIVLPS 544
Db 658 ATKPENSNGNTSNGQTEPEKLELN--VSDIELYSQ--TNGT---YRQVSLDGIPE 710
QY 545 VQ-----EIGRSF-----RONGAN--NLIFMGSKVKTIGEMAFLSNRLEHLD 585
Db 711 TDTYFVKVKSAPKQVYIPVASITEKRGQSVYKITAKAEKLQBELENKYVDNFTFYLD 770
QY 586 LSEKQOLTEIPVOAFSDNALKEVLLPASLKTIRREAFKKNHLKQLEVASALSHIAPNALD 645
Db 771 KKAKENTNF--TSFS-NLVK-----AINQN-----PSGTHLA-ASLN 805
QY 646 DNDGBEQFNKVVVTHNSYALA--DGEHFVDPDKLSTTIVLEKILKLEGLDYSTL 703
Db 806 ANEVELSTDCKSVIKGFTGQIGEKDGKHY-----AIYNLKK--PLFENLSGATV 854
QY 704 RQTQTQFQDMWTAGKALLSK--SNLRQEKQKFLQEAQFFLGRVDLDKAIKAEKALVT 761
Db 855 EKLS---LKNVAISGRNDIGSLANEATNGTKIKQVHVDGVLAGEGVRGGLLAKAQSSIA 911
QY 762 KKATKNGQLLERSIN--KAVLAYN-----NSAIKK-----ANVKRLEKELD 800
Db 912 ESFPKG-----RIVNTYETDTSNIGGLVGHLTGKNASIAKSAKATVITISNTNRSDQTVG 966
QY 801 LUTGLVEGKPLAQATMVGCVLLKTPPLPEYIIGL-----NVYFDKSGKLIYALDMS 854
Db 967 GLAGLVDRDAQIQDSVAEGDINNPK-----HFGRVAGVAGNLWDRTSGDVVRHAGSLT 1018
QY 855 DTIGEQCKDAYGNPILNVDBDN--EGYHALAVATLADYEGLDIKITLNSKLSQTSI--- 909
Db 1019 NVLSD-----VNVNNGNAITGYH-----YTGKMKVANTFSSKANRVFNVTLE 1059
QY 910 --RQVPTAAVHRAGIFQAIONAAAEQQLPKPGTHS-----EKSSSESANSKD 957
Db 1060 KDVVSKESEFERG-----TWLDASQIVSKAEISPLTLPVPELSTSGKKSDFSKI 1112
QY 958 RGLQSNPKTNRGRHSAILPRTGSGSVFYGILGYTISVALL--SLITAIKKK 1006
Db 1113 AHVQANRALVYKNIEKLLFPYKNSTIVKYGNLVKNSLIYQKELLSAVNMK 1163
```

RESULT 34

US-09-107-433-3251

; Sequence 3251, Application US/09107433

```
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bueh
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
;
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 3251:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1972 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
;
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1972
; SEQUENCE DESCRIPTION: SEQ ID NO: 3251:
;
; US-09-107-433-3251
```

```
Query Match 3.3%; Score 169.5; DB 4; Length 1972;
Best Local Similarity 19.4%; Pred. No. 0.008;
Matches 231; Conservative 167; Mismatches 442; Indels 351; Gaps 56;

QY 4 HLKTVALLTITVSVVTH-----NOEVFSLVKEPIKQTCASSISGADYA--- 48
Db 143 HLVLVAGIGNGLVSVHLENHLLNNTDYELTSGEKLPLPK-----ISGYTVIGY 196
QY 49 -----ESSGKSKLKINET-SGPVDDTVTDLFSDKRTPPEKID-----N 86
Db 197 KEGKTTSDFEVSNQESAAATPTKQKQVDVNTVNFVDHPSTVQAIEQTPVSSTKPTEVQ 256
QY 87 LAKGPKEQL-----KAVTTESEKQITSGSOLESKESLSLNTKVPSTSNWEICDFITK 142
Db 257 VVEKFPSTELINPRKEEQSSDSQQLAEHKNLETKKE-----EKISPK----- 300
QY 143 GNTLVGLSKSGVEKLSQTDHLVLPQQAADGTOLIQVAFPTDPKKTAIAEYTS----- 196
Db 301 -----EKTGVNTLNPQDE-VLSGQLNKPPELLYREETIEIKIDFQEEIQENPDIVGIV 352
QY 197 RAGENGESQLDVGKEI-----INGEVFNYSLLKKVTIP----- 232
Db 353 RVKQEGKL-----GKKVEIRVIFSVNKEBVSREIVSTSTTAPRIVEKGTGKTQVIKE 406
```

```
QY 233 ---TCYKHIG-----QDAFVDMKNIAEVLNLPESLETISDYAPAHALKOI 274
Db 407 QPETOGEHKDVOSGAIVPEPAIOPELPEAVVSDKGPEVOPALSKAVITDKGETEV---QP 463
QY 275 DLPDNLKAIGELAFPDNQITGKLSLQRLMLAERAFKSNHIKTIEFRGNSLKVIGRASF 334
Db 464 BSPDT-----VVSCKGPEQVAPLP-----EYKGNIEQVKPPTPV 498
QY 335 QNDLSQLMPLDGLKIB-----SEAFGNPGDDHNNRVVLTWKSGNPSGLATENT- 387
Db 499 EKT-----EQPEKTEBPVKPTEETPVNNEGTEGTSI---QEAENPVQAEESTT 549
QY 388 ---VYNPDKS---LWQES-----PEIDYTKMLEEDFTYQKNSVTGFSNKG 426
Db 550 NSEKVPSPDTSSENTGEESNPSDSTTSVGESKNKPHNSKNSSEKTEVEEVPVN---PNEG 607
QY 427 LQKVRNKNLEIPKQHGVTIETIGDNAFRNV-DPQNK-TLRKYDLEEVKLPSSTIRKICA 484
Db 608 TVEGTSNQETEKPVQPAEETQNSCKIANENTGEVSNKPSDSKPPVEESNQ---EKNGT 664
QY 485 FAFQNNLKSFRASDDLEBEIKGAPMNRRIETLEKDKLVITGDAAFHINHIYAILVLES 544
Db 665 ATKPENSNGNTSNGQTSPEKKLELRN--VSDIELYSQ--TNGT---YRQHVSLDGIPEN 717
QY 545 VQ-----EIGRSF-----RONGAN--NLIFMGSKVKTGEMAFLSNLEHL 585
Db 718 TDTYFVKVKSAPKDVYIPVASITBEKNGQSVYKITAKAEKLOELENKYVDNFTFYLD 777
QY 586 LSEKQLTEIPVQAFSDNALKEVLLPASLKTIRBEAFKKNHLKQLEVASALSASHIAFNALD 645
Db 778 KKAKEENTNF--TSFS-NLVK-----AINQN-----PSGYHLA-ASLN 812
QY 646 DNDGEQFQDNKVVVTHNSYALA--DGEHFIVDPDKLSSTIVDLEKILKLEGLDYSTL 703
Db 813 ANEVELSTDDKSYIKGTFTGQLIGSKDGHY-----AIYNLKK--PLPENLSGATV 861
QY 704 RQTOTQFDMTATGALLSK--SNLROGEKQKFLQEAQFFLGRVYDLDKATAKAELAVT 761
Db 862 EKLS---LKNVAISGNDIGSILANBATNGTKIKQVHVGVLAGERGVOGLLAKADQSSIA 918
QY 762 KKAATNGQLLERSIN--KAVLAYN-----NSAIKK-----ANVKLEKELD 800
Db 919 ESSFKG---RIVNTYETDTSYNGVLGHLTGKNSIAKSAKATVITSSNTNRSQIVG 973
QY 801 LITGLVEKGPLQAQTMQGVYLLKTPILPLPYIYIGL-----NVYFDKSGKLIYALDMS 854
Db 974 GLAGLVDRDAQIQDSYAEGDINNPK-----HFGRVAGVAGNLWDRTSQDVRHAGSLT 1025
QY 855 DTIGEGQKDAYGNPILNVDEDN--EGYHALAVATLADYEGLDIKILNSKLSQTSI--- 909
Db 1026 NVLSD-----VNVNTGNAITGYH-----YTGMMKVANTFSSKANRVFNVTLE 1066
QY 910 --RQVPTAAYHRAGIFQAIQNAAEABQLPKPGTHS-----EKSSESSEANGKD 957
Db 1067 KDEVVSKESEFBERG-----TMLDASQIVSKAHSPLTLPTVEPLSTSGKSDSFSKI 1119
QY 958 RGLQNPKNRGRHSAIPLRTGSKGSFVYILGYTSVALL--SLITAIKK 1006
Db 1120 AHYQANRALVYKNIEKLLPFYKNSTIVKYGNLVKNSLLYQKELLSAVMMK 1170
```

## RESULT 35

```
US-09-107-532A-3855
; Sequence 3855, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
```

```
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arianiello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3855:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1095 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...1095
; SEQUENCE DESCRIPTION: SEQ ID NO: 3855:
US-09-107-532A-3855
```

```
Query Match 3.3%; Score 169; DB 4; Length 1095;
Best Local Similarity 19.4%; Pred. No. 0.0036;
Matches 231; Conservative 171; Mismatches 431; Indels 358; Gaps 62;

QY 8 VALILTVSVVTHNQEV-----FSLVKEPILKQTOASSISGADYAEBSG-KSKLKIN 59
Db 46 IAPTITFADEITHPTQTVTVELDLAHQYVVEGTFSDGRPMSEVTPHYAVYNGVKQDIFCI 105
QY 60 ETSQGPVDDTVTDLFSDKRTTPEKIKDNLAKGPREQELKAV---TENTSEKQITSGSOLE 116
Db 106 EPGVPIDN-----EFTPGYKPNLPDMPERKALVSLWKAGTDVDTTHVAQKMIW 156
QY 117 QSKESLSLANK-----TVPSTSNMWCIDFITK-----GNTLVGLSKSGVE-- 155
Db 157 QEVNGYTLHSIKRLNGSAVNIAAEAKINQAIADYQKKPSFHNSTAKIVLGSQTTMTDTN 216
QY 156 --KLSQTDHLVLPQAAD-----GTQLIQVASFATPD---KKTAAIETVSAGENGELISQ 206
Db 217 NLNLSEFDEVVENTANIDYRVNGNQLI-----ITPNASSKESGVLTKKSAGTGPVAY 270
QY 207 LDVPGKEILNEG-----EVFNSYLLKKV-----TIPTGYKHI-----G 239
Db 271 KKV-QGQTLMAGAIKDKNTYTVTKIDVTEGSLKIKKVDKESGALVPGTVFHLDFGKNLPA 329
QY 240 QDAFVDMKNIAEVLN-LPESLE-TISDYAF-AHLAKQIDLPDNLKAIGELAFPDN---QI 293
Db 330 KDVITDKEGIATLDEIPHGTKTIVITEKSVAPYPTIDTTPMTTIKA-GETIVVTSKNAQE 388
QY 294 TGLKLSLQRLMLAERAFKSNHIKTIEFRGNSL-----KVIGEASFQNDLSQL-- 342
Db 389 KGQIILDKSGVETGSDMLNDNY-----SLAGNTFAIRKDSPTGIEIVQEMTTDENGHAETPK 444
QY 343 MLPDGLE-----KIESEAFTCNPGD-----DHYNNRVVLTWKSGNPSGLAT-ENTY 388
```





Db 1044 SGEAFKLKSLR-----VLMNANNLQSLPABISQL-----SQLSVIDVG 1083  
Qy 777 KAVLAYN-----NSAIKANVK---RLEKELDLTLGLVEGKPLAQATWQGV 821  
Db 1084 SNQKYNISNYHDWNRQNTLKYLFNFSGNTRPEIKSAMGYGTMMHLSDLTVLKQLRVL 1143  
Qy 822 YLLKTPPL---PLPEYIYGLNVYFDKSGKLIIVALDMSDTIGEGQDAYGNPIL-----N 871  
Db 1144 GLMDVTLNTSRVPD---DGNVFLRLTNGSTINGMEYGVADSLQDTDSVTRDTPFRFGN 1201  
Qy 872 VDEDEGYHALAVATLADYEGLDIKTLNLSKLSQLTS-----IRQV-----PTAAAYHR 919  
Db 1202 EDE-----CLICLYDGINDNTNGHKISQVIRDIYDRILVRSVEKYGDSTPABIR 1251  
Qy 920 AGI-FOATONAAAEQQLPKPGTHSEKSSSESANSKDRGLQSPKTNRGHSAILPRT 978  
Db 1252 DALREFSLQNLKNEINSSILAAGNNKNSNGNTNNNNNTNNNNNGNNTNNGESVYLTAA 1311  
Qy 979 GS-----KGSFVYGLGYTSVALLS 998  
Db 1312 DLSGASVTVPFMKGNITNIGNATALLS 1342

## RESULT 37

US-08-446-137B-2  
Sequence 2, Application US/08446137B

Patent No. 6162903

GENERAL INFORMATION:

APPLICANT: Trowern, Angus R.

APPLICANT: Ackinson, Anthony

APPLICANT: Murphy, Jonathan P.

APPLICANT: Laurence, Oliver S.

APPLICANT: Duggleby, Clive J.

TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED

TITLE OF INVENTION: FROM L PROTEIN AND THEIR USES

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSES: SEED and BERRY LLP

STREET: 6100 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/446,137B

FILING DATE: 22-MAY-1995

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: McMaster, David D.

REGISTRATION NUMBER: 33,963

REFERENCE/DOCKET NUMBER: 100084.406

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1027 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-446-137B-2

Query Match 3.3%; Score 168.5; DB 3; Length 1027;

Best Local Similarity 19.2%; Pred. No. 0.0036; Indels 323; Gaps 56;

Matches 221; Conservative 174; Mismatches 436;

Qy 20 HNQEVFSLVKEPILKQTOASSISGADYAESSGSKLKIINETSQPVDDTVTDLFSDKRTT 79

Db 31 HCQNMKINKKLLMAALAGAIUVGGGANAYAAERDNTDNNLSMDDEISDAYPDYHGD---- 86  
Qy 80 PEKIKDNLAKPRQEL-----KAVTENTE-SEKQITSGSQLEQSKESLSLNTKVPSTSN 133  
Db 87 ---VSDSV--DPVBEEDIALAKALAEAKETAKKHIDSLNHLSETAKKLAKNDIDSATTI 141  
Qy 134 WEICDFTTKGNTLVGLSKSGVEKLSQTDHLVLPQQAADGTQLIQVASFAPDPDKKTAIAE 193  
Db 142 NAINDIVARADV---MERKTAEK-EAEKLAATAETA-----KXHIDE 180  
Qy 194 YTSRAGENGEISQDVGKEIINEGEVFNFSYLLKKVTIPTGYGKHIGQAFVNDKNIAFWN 253  
Db 181 LKHLADTKELAKRIDSAATTIN---AINDIVAR-----ADVMERKTAKE 223  
Qy 254 LPESLETISDYAFAH-----LALKQIDLPDNLKAIGELAPFDNQITGKLS-- 298  
Db 224 EAEKLAATAETAKKHIDELKHLADTKELAKRIDSAATTINDIINVARADVMERKLUSEK 283  
Qy 299 ---LPROLMRL-ABRAPKSNHIKTIEFRGNSLKVIGEASFQDNDSLQMLPDGLEKIBSE 354  
Db 284 ETPPEBEVTIKANLIFADGSTQNAEFKGTFAKAVSDA-----YAYADALKXONGE 334  
Qy 355 AFTGNPGD-----DHYNNRVLMTKSGKNPSG-----LATENT 387  
Db 335 -YTVDVADKGLTLMIKPAGKKEPEEPKEEVTIKVNLIIFADGKTQTAEFKGTPEEATAKA 393  
Qy 388 YVNPDKSLWQESPEIDYTKMLBEDFTYQKNSVT-GPSNKGLOKVRKNKLIIPQHNGVTV 446  
Db 394 YAYAD-LLAKENGE--YTADLEDG---GNTINIKPAGK-----ETPETPEEPKEEVTIK 441  
Qy 447 ITEI-GDNAPRVDFONKTLRKYLEEVKLPSTIRKIGAPAFQSNLKLSPFASDDLEBIK 505  
Db 442 VNLIFADGKIQTAEFK-----GTPEEATAKAYAYANLLAKENGEYTDLE 486  
Qy 506 EGAPMNNRI-----ETLELKDKLVTITGDAAPHNHIIYAILVPESVQEIQRSAFRQ 555  
Db 487 DG---GNTINIKPAGKETPETPEEPKEEVTI---KVNLIIFADGKTQTAEFKG--TPEE 536  
Qy 556 NGANNLIFPMGSKVKTLGEMAFLSNRLE-----HLDLSEQQLTEIPQAFSDNALKEVL 609  
Db 537 ATAERYADLLAKVNGE---YTADLEDGGYTTINIKPAGKEQPGENP-----GITIDSBL 588  
Qy 610 LPASLKTIRBEAFKKNHLKQLEVASALSASHIAFNALDNDGDQFDNKKVVVKT--HNSYA 667  
Db 589 ----LKAKEEAIKE--LKEAGITSDLYFSLINKAKTVEGVEALKNE-ILKAHAGEETPE 641  
Qy 668 LADGEHPFIVDPDKLSSTIV---DLEKILKLIEGLD---YSTLRQTTQTQFRDMTTAGKAL 721  
Db 642 LKDGATVYEEAABAAAKEALKNDVDVNNAYEIVQADGRYYVVLK-----I 685  
Qy 722 LSKSNLRQGEKQKFLQEAQFFLGRVLDLKATAKAEKALVTYKATKNGGOLLERSINKAVLA 781  
Db 686 EVADSEEFGEDETPVEQE---GYATYEEAABAAAKEA-----LKEDKVNNAVEV 729  
Qy 782 YNNSAIKKANVKRLEKELDLTLGLVEGKPLAQATWQGV---YLLKTP-----LPLPE 832  
Db 730 VQADGRYYVYVLKIEDKEDSQPGEEGENP-----GITIDEWLLKNKAKDAIKELKE 781  
Qy 833 YYIGLVNYFD--KSGKLIYALD-MSDTIGEGQKDAYG-NPILANVDE----- 874  
Db 782 AGISSDIYFDAINKAKTVEGVEALKNEILKAHAEKPGENPGITIDEWLLKNKAAAIKEL 841  
Qy 875 -----DNEGYHALAVATL---ADYEG-----LDIKTLNLSKLSQLT 907  
Db 842 KEAGITAEYLFNLINKAKTVEGVESLKNELKAHAEKPGENPGITIDEWLLKNKAKDAIK 901  
Qy 908 SIRO--VPTAAVHRA-----GIFOAIQNAAEAEQLLPKPG-----THSEKSSS 949  
Db 902 ELKEAGITSDIYFDAINKAKTIEGV-EALKNEILKAHKKDEEPKCKPGEDKKPKDKE 960  
Qy 950 SESANSKDRGLQSNP-----KTNRG-----RHSAILPRTGSKGSFVYIGLVTSVALLSL 999  
Db 961 DKPKEDKPKGEDKKPKDGTOKDSPNKKKAKLPAKAGSEAEIL-----TLAAAL 1013



```
QY 169 AADGTQLIQVA--SFAETPD-----KKTATAEYTSRAGENCEISQLDVDGKEIINEGEVENS 223
Db 169 AADGTQLIQVA--SFAETPD-----KKTATAEYTSRAGENCEISQLDVDGKEIINEGEVENS 223
QY 52 AKDGKKDSAAQPGVAFSDVNTIKRPNPAGTRKKSNAEVI-----KB-LNKCREENS 104
Db 52 AKDGKKDSAAQPGVAFSDVNTIKRPNPAGTRKKSNAEVI-----KB-LNKCREENS 104
QY 224 YL--LKKVTIPRGYKHIGQADAVDNKNTAEVNLPSLETISDYAFALHAKQI-DLPDN- 279
Db 224 YL--LKKVTIPRGYKHIGQADAVDNKNTAEVNLPSLETISDYAFALHAKQI-DLPDN- 279
QY 105 MRLDLSKRSI-----HI-----LPSIKELTQLTLYLYSNKQLSLPAEV 144
Db 105 MRLDLSKRSI-----HI-----LPSIKELTQLTLYLYSNKQLSLPAEV 144
QY 280 --LKAIGELAFPDNQTGKLSLPROMLRAEAFKSNHIKTIFFRGNLSKLVIGEASFDN 337
Db 280 --LKAIGELAFPDNQTGKLSLPROMLRAEAFKSNHIKTIFFRGNLSKLVIGEASFDN 337
QY 145 GCLVNLMTLALSNSLT--SLPDSLDNLK-----LRMLDLRHNKLRITPSVVRDL 194
Db 145 GCLVNLMTLALSNSLT--SLPDSLDNLK-----LRMLDLRHNKLRITPSVVRDL 194
QY 338 DLSQMLPDGLEKIESEAFTPGDDHYNNRVLMTKSGKQPSGLATENTYVNPDKSLWQ 397
Db 338 DLSQMLPDGLEKIESEAFTPGDDHYNNRVLMTKSGKQPSGLATENTYVNPDKSLWQ 397
QY 195 SLTTLYL-----RPNRTTVEKDKIKLSLMSIRENKIKQLPA 234
Db 195 SLTTLYL-----RPNRTTVEKDKIKLSLMSIRENKIKQLPA 234
QY 398 ESPEIDYTKWLEEDFTYQKNSVTGFSNKGQLQVKNKLEIPKQHGVTITEIGD-NAPR 456
Db 398 ESPEIDYTKWLEEDFTYQKNSVTGFSNKGQLQVKNKLEIPKQHGVTITEIGD-NAPR 456
QY 235 EIGEL-----CNLIT-----LDVAHQLEHLPK-----EIGNCTQIT 266
Db 235 EIGEL-----CNLIT-----LDVAHQLEHLPK-----EIGNCTQIT 266
QY 457 NVDFQNKTKRYDLEEVKLPSIRKIGAPA-----FOSNML-----KSFASDDLLEIKGAP 509
Db 457 NVDFQNKTKRYDLEEVKLPSIRKIGAPA-----FOSNML-----KSFASDDLLEIKGAP 509
QY 267 NLDLQHNEI-----LDLPDTIGNLSLSRLGLRYNRLSAIPRSLAKCSALEELN---L 316
Db 267 NLDLQHNEI-----LDLPDTIGNLSLSRLGLRYNRLSAIPRSLAKCSALEELN---L 316
QY 510 MNRNRIETL--ELKDKLVITIGDAAPHIN--HIYAIVLPSVQVQIGRSAFRQNGANNLIF-M 564
Db 510 MNRNRIETL--ELKDKLVITIGDAAPHIN--HIYAIVLPSVQVQIGRSAFRQNGANNLIF-M 564
QY 317 ENNNISTLPESLSSLVKLNLSUTLARNCFQLYPVGGPQSFSTIYSLNMEHNRINKIPGI 376
Db 317 ENNNISTLPESLSSLVKLNLSUTLARNCFQLYPVGGPQSFSTIYSLNMEHNRINKIPGI 376
QY 565 GSKVKTLAGEMAFSLNRLHLDLS-----EQQLTEIP-----EQAFSD 602
Db 565 GSKVKTLAGEMAFSLNRLHLDLS-----EQQLTEIP-----EQAFSD 602
QY 377 FBRKVLKLNKMDKQLSLPLDFGTWTSWVFNLTATQTKLPEDVSLVSLVILSN 436
Db 377 FBRKVLKLNKMDKQLSLPLDFGTWTSWVFNLTATQTKLPEDVSLVSLVILSN 436
QY 603 NALKEVLPL--ASLKTRIBREAFKQHLKQL--EVA 633
Db 603 NALKEVLPL--ASLKTRIBREAFKQHLKQL--EVA 633
QY 437 NLLKK--LPHGLGNLRKRLRELDLEENKLESIPNEIA 470
Db 437 NLLKK--LPHGLGNLRKRLRELDLEENKLESIPNEIA 470
```

## RESULT 40

```
US-08-053-614-2
; Sequence 2, Application US/08053614
; Patent No. 5403924
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: TUMURU, MURALI K.R.
; TITLE OF INVENTION: THE taga GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSES: NEEDLE & ROSENBERG, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/053,614
; FILING DATE: 19930426
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 859 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-053-614-2

Query Match 3.3%; Score 167.5; DB 1; Length 859;
Best Local Similarity 20.4%; Pred. No. 0.0032;
Matches 212; Conservative 150; Mismatches 344; Indels 331; Gaps 59;

QY 27 LVKEPILKQTQASSISGADYAESGSKLKINETSGVDVTDVTLDFSKDKRTTPEKIDN 86
Db 27 LVKEPILKQTQASSISGADYAESGSKLKINETSGVDVTDVTLDFSKDKRTTPEKIDN 86
QY 1 MTNETIQDPQTEAFAFPQPFNNLQVAFK-----VDNAVASYDPDQKPIVDK---- 49
Db 1 MTNETIQDPQTEAFAFPQPFNNLQVAFK-----VDNAVASYDPDQKPIVDK---- 49
QY 87 LAKGPREQELKAVTENTESEKQITSGSQLESKLSLNTKV--PSTSNWEICPITKGN 144
Db 87 LAKGPREQELKAVTENTESEKQITSGSQLESKLSLNTKV--PSTSNWEICPITKGN 144
QY 50 -----NDRDNRQAFEG--ISQLREYS-NKAIKNPTKQNFYDFINKSN 91
Db 50 -----NDRDNRQAFEG--ISQLREYS-NKAIKNPTKQNFYDFINKSN 91
QY 145 TLVG-----LSKSGVEKLSQTDHLVLPQAADGTQLIQVAFAP-----TP 185
Db 145 TLVG-----LSKSGVEKLSQTDHLVLPQAADGTQLIQVAFAP-----TP 185
QY 92 DLINKMDLIVVESSTKSFQKPGDQRYRIFTSWVSHQNDPSKINTRCIRNFMEHTIQPIP 151
Db 92 DLINKMDLIVVESSTKSFQKPGDQRYRIFTSWVSHQNDPSKINTRCIRNFMEHTIQPIP 151
QY 186 DKTAIAEYTSRAGEN--GEI--SQLDVGKEIINEGEVFNLSYLLKKVTI-----PTGYK 236
Db 186 DKTAIAEYTSRAGEN--GEI--SQLDVGKEIINEGEVFNLSYLLKKVTI-----PTGYK 236
QY 152 DDEKE-APFLKSAQSPAGIIGNQIRTDQFM---GVFDES LKERQAEKNGPTGD 206
Db 152 DDEKE-APFLKSAQSPAGIIGNQIRTDQFM---GVFDES LKERQAEKNGPTGD 206
QY 237 HIGQDAFV---DNKNIAEAVNLPSLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQ 292
Db 237 HIGQDAFV---DNKNIAEAVNLPSLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQ 292
QY 207 WL--DIFLSFDFKQSSDVK-----EAINQEPLPHVQ-----PD-----IATSTTH 246
Db 207 WL--DIFLSFDFKQSSDVK-----EAINQEPLPHVQ-----PD-----IATSTTH 246
QY 293 ITGKLSLPROMLRAEBA--PKSNHIKTIETPRG-----NSLKVIGEASFDND 338
Db 293 ITGKLSLPROMLRAEBA--PKSNHIKTIETPRG-----NSLKVIGEASFDND 338
QY 247 IOGLPPESRDJ--LDERGNFSKFTGLDMEMLDVGEVADMDPNYKFNQLLI-----HNNT 298
Db 247 IOGLPPESRDJ--LDERGNFSKFTGLDMEMLDVGEVADMDPNYKFNQLLI-----HNNT 298
QY 339 LSQMLM--PDGLEKIE--SEAFTPG-----DDHYNNRVLMTKSGKQPSG 381
Db 339 LSQMLM--PDGLEKIE--SEAFTPG-----DDHYNNRVLMTKSGKQPSG 381
QY 299 LSSVLMSHSDGIEPEKVSLLYAGNGFGAKHDNATVGYKQOQGNVATIINVHMKSG 358
Db 299 LSSVLMSHSDGIEPEKVSLLYAGNGFGAKHDNATVGYKQOQGNVATIINVHMKSG 358
QY 382 LAT---ENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGQLQVKNK--NL 436
Db 382 LAT---ENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGQLQVKNK--NL 436
QY 359 LVIAGEGKGINP-----SPLYKEDQLTGSQALSQBEIQNKIDFM 400
Db 359 LVIAGEGKGINP-----SPLYKEDQLTGSQALSQBEIQNKIDFM 400
QY 437 EIPQHNGV--TITEIGNAPRN--VDPONKTKRYDLEEVKLPSIRKIG--AFASQSN 490
Db 437 EIPQHNGV--TITEIGNAPRN--VDPONKTKRYDLEEVKLPSIRKIG--AFASQSN 490
QY 401 EFLAQNNAKLDLSSEKEKEFRNEIKDFQ--KDSKPY-----LDALGNDRIFAVSK 449
Db 401 EFLAQNNAKLDLSSEKEKEFRNEIKDFQ--KDSKPY-----LDALGNDRIFAVSK 449
QY 491 NLKSFEASDDLLEIKEG-----AFMNRNRIETLEKDKLVITIGDAAPHINHIYAIVLPE 543
Db 491 NLKSFEASDDLLEIKEG-----AFMNRNRIETLEKDKLVITIGDAAPHINHIYAIVLPE 543
QY 450 --KDPKHSALITEPNKGDLSYTLKVMGKQIKALD-REKNVTLOGNLKH-----D 496
Db 450 --KDPKHSALITEPNKGDLSYTLKVMGKQIKALD-REKNVTLOGNLKH-----D 496
QY 544 SVQBIGRSAPRQNGANNLIFMGSKVKTGEMAFPSNRLHLDLSEQ-----Q 591
Db 544 SVQBIGRSAPRQNGANNLIFMGSKVKTGEMAFPSNRLHLDLSEQ-----Q 591
QY 497 GVMFVNTSNPKYTNASK-----SPNKGVG-----VTNGVSHLEAGFSKAVFNLPLNLA 547
Db 497 GVMFVNTSNPKYTNASK-----SPNKGVG-----VTNGVSHLEAGFSKAVFNLPLNLA 547
QY 592 LTEIPVQAFPSNALKEVLLPASLKTIRBEAFKQHLKQLEVASALSAPNALDDNDGDE 651
Db 592 LTEIPVQAFPSNALKEVLLPASLKTIRBEAFKQHLKQLEVASALSAPNALDDNDGDE 651
QY 548 ITSVRRDLEDKLIAGLSQAEANKLVKDFLSSN--XEL-VGKALNF----- 591
Db 548 ITSVRRDLEDKLIAGLSQAEANKLVKDFLSSN--XEL-VGKALNF----- 591
QY 652 QFDNKVVVKTTHNSYALADGEHFIVDPDKLSSTTVDLEKILKLEGLDYSLRQTQTQF 711
Db 652 QFDNKVVVKTTHNSYALADGEHFIVDPDKLSSTTVDLEKILKLEGLDYSLRQTQTQF 711
QY 592 ---NKVAEAKNTG-----NYDEVKRAQDKLEKSLKKEHLE----- 625
Db 592 ---NKVAEAKNTG-----NYDEVKRAQDKLEKSLKKEHLE----- 625
QY 712 RDMTTAGKALLSKNLRQGEKQKFLQBAQFFLGRVLDLDAJAKAEK-----ALVTYKATPN 767
Db 712 RDMTTAGKALLSKNLRQGEKQKFLQBAQFFLGRVLDLDAJAKAEK-----ALVTYKATPN 767
QY 626 KDV---AKNLESKS---GNKNKMEAKAQ-----ANSQKOEIFALINKENRD 666
Db 626 KDV---AKNLESKS---GNKNKMEAKAQ-----ANSQKOEIFALINKENRD 666
QY 768 GQLLE-----RSINKAVLAYNNSAI-----KKNVKELEKELDLTLGLV 806
Db 768 GQLLE-----RSINKAVLAYNNSAI-----KKNVKELEKELDLTLGLV 806
QY 667 ARAIYAYQNKGIRKRELSKLENINKOLKDYFSKDFGFKNGKNDKFSKAEETLKALKGSV 726
Db 667 ARAIYAYQNKGIRKRELSKLENINKOLKDYFSKDFGFKNGKNDKFSKAEETLKALKGSV 726
QY 807 EGKGPLAQATWVGQVYLLKTP---LPLPEYVIGLVNVPDKSGKLIYALDMDSDTIGEGKD 863
Db 807 EGKGPLAQATWVGQVYLLKTP---LPLPEYVIGLVNVPDKSGKLIYALDMDSDTIGEGKD 863
QY 727 KDLG-----INPEWISKVENLNAALNEFKNGKNDKFSKVTQA-----KSD-----QEN 769
Db 727 KDLG-----INPEWISKVENLNAALNEFKNGKNDKFSKVTQA-----KSD-----QEN 769
```



STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/034,306  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,397  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: SPRATT, GWENDOLYN D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 2200.030  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 859 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-034-306-2

Query Match 3.3%; Score 167.5; DB 2; Length 859;  
Best Local Similarity 20.4%; Pred. No. 0.0032;  
Matches 212; Conservative 150; Mismatches 344; Indels 331; Gaps 59;

QY 27 LVKEPILKQTAASSISGADYAESGSKKINETSQGVDDVTVDLFSDKRTTPKIKDN 86  
DB 1 MTNETIDQPTQEAAPNQQFINNQLVAPLK-----VDNAVSYPDPQPIVDK---- 49  
QY 87 LAKGPRBELKAVTENTSEKQITSGSQLEQSKESLSLNTKV--PSTSNWEICDITIGN 144  
DB 50 -----NDRDNROAFEG--ISQUREEYS-NKAIKNPTKKNQYFSDPINKSN 91  
QY 145 TLVG-----LSKSGVEKLSQTHVLPLPSQAADGTQLIQVAFAP-----TP 185  
DB 92 DLINKONLIVSESSYTSKQFQDQYRIFTSWVSHQNDPSKINTCIRNFHEHTQPPIP 151  
QY 186 DKTAIAEYTSRAGEN--GEI--SOLDVDGKEIINEGEVFNLSYLLKQVTI-----PTGYK 236  
DB 152 DKKEK-AEFLKSAKQSFAGIIGNQIRTDQKPM---GVFDSLSKERQEAENKGGTGD 206  
QY 237 HIGQDAFV---DNKNIAEVNLPESLETISDYAFAPHLAKQIDLDPNLKAGELAFFNQ 292  
DB 207 WL--DIFLSFIDFKQSSDVK-----EAINQBPPLHVQ-----IATSTH 246  
QY 293 ITGKLSLPLQLMLAERA-----FKSNHIKTIIFRG-----NSLKVGEASFQND 338  
DB 247 IQLPPESSDL--LBERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLI-----HNNT 298  
QY 339 LSQML--PDGLEKIE--SEAFGNPQ-----DDHYNNRVVLWTKSGKNPSG 381  
DB 299 LSSVLGSHDIEPEKVSLLYAGNGFGAKHDWNAVGVYKQDQGNVATIINVHMGSG 358  
QY 382 LAT---ENTYVNPDKSLWQESPEIDYTKWLEBDFYQKNSVTGFSNKGLOKVRNK--NL 436  
DB 359 LVIAAGEKGINP-----SFYLYKEDQLTGSQRALSQEBIQNKIDPM 400  
QY 437 EIPKQHNGV--TITEIGNAFEN--VDFQNTLRLKYDLEEVKLPTIRKIG--AFAPQSN 490  
DB 401 EFLAQNAKLDSLSKEKEKFRNETKQF--KDSKY-----LDALGNDRIFAFVK 449  
QY 491 NLKSPFASDLEIEKEG-----AFMNNRIETLEKDKLVITGDAAFHINHIYAILVPE 543  
DB 450 --KDPKHSALITFENKGDLSYTLKVMGKKQIKALD-REKNVTLQGNLKH-----D 496

QY 544 SVQEIGBSAPRQNGANNLIIFMGSKVKTIGEMAPLSNRLEHLDLSEQK-----Q 591  
DB 497 GVMFVNTYSNFKYTNASK-----SPNKGVG-----VTNGVSHLEAGFSKVAVFNLPNLNLMA 547  
QY 592 LTEIPVQAFSDNALKEVLLPASLKTIREEAPKQNHKLQLEVASALSASHIAFNALDNDGDE 651  
DB 548 ITSVVRLDEKLIAGLSPOEANKLVKDFLSSN--KEL-VGKALNP-----591  
QY 652 QFDNKVVVTHNSYALADGEHFIVDPDKLSSTIVDLKLSLIEGLDYSTLRTTOTQF 711  
DB 592 ---NKAVAEAKNTG-----NYDEVKRAQKOLEKSLKREHLE-----625  
QY 712 RDMTTAGKALLSKSNLRQGEKQKFLQEAQPLGRVLDLDAKAKAEK---ALVTKKATKN 767  
DB 626 KDV---AKNLESKS-----GNQNMKAQAQ-----ANSQOEIPALINKENRND 666  
QY 768 QOLLE-----RSINKAVLAYNNSAI-----KCANVVRLEKELDLTLGLV 806  
DB 667 ARAIYAQNLKGIKRELSDKLENINKDKLQKFSKDFGKNGKNDKDFSKAEETLKALKGSV 726  
QY 807 EGKGLAQAQWVGYYLLKTP---LPLPEYYIGLVNVPYFDKSGKLIYALDMSDTTIGEGKD 863  
DB 727 KDLG-----INPEWISKVENLNAALNEPKNGKNDKFSKVTQA---KSD-----QEN 769  
QY 864 AYGNPILN-----VDEDNEGYHALAVATLA-DYEGL-----DIKTILNSKLSQTSIRQ 911  
DB 770 SIKDVIINQKITDKVDELNQ---AVSVAKTACDPSGVEQALADLKNFSKEQLAQ--QAOK 824  
QY 912 VPTAAYHRAGIFQAIQN 928  
DB 825 NESFVNGKSEIYQSVKN 841

RESULT 43  
US-09-259-437-2  
; Sequence 2, Application US/09259437  
; Patent No. 6153390  
; GENERAL INFORMATION:  
; APPLICANT: COVER, TIMOTHY L.  
; APPLICANT: BLASER, MARTIN J.  
; APPLICANT: HARRY KLEANTHOUS  
; APPLICANT: TUMURU, MURALI K.R.  
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING  
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
; STREET: 127 Peachtree Street, Suite 1200  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/259,437  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/316,397  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SPRATT, GWENDOLYN D.  
; REGISTRATION NUMBER: 36,016  
; REFERENCE/DOCKET NUMBER: 2200.030  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404/688-0770  
; TELEFAX: 404/688-9880  
; INFORMATION FOR SEQ ID NO: 2:

### SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:  
LENGTH: 859 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

Query Match 3.3%; Score 167.5; DB 3; Length 859;  
Best Local Similarity 20.4%; Pred. No. 0.0032;  
Matches 212; Conservative 150; Mismatches 344; Indels 331

Qy	27	LVEPIILKQTOASSISGADYAESGSKL KINETS GPVDVTVDL FSDKRTTEPEKIKDN	86
Db	1	MTNETIDQOPQTEAAFPNQQPINNLVAFLK-----VDNAVASYDQDPKPIVDK	49
Qy	87	LAKGPREQELKAVTENTESEKQITSGSOLESKESLSLANKTV--PSTSNWEICDPITKGN	144
Db	50	-----NDRDNQAPEG-----ISOURBEYS-NKAIKNPTKKNQNYFSDFINKSN	91
Qy	145	TLVG-----LSKSGVEKLSQTDHLVLPQQAADGTQLIOVASFAF-----TP	185
Db	92	DLINKONLIVVESTKSFQKFGDQRYRIFTSVWSHQNDPDKINTFCIRPNFMEHTIOPIP	151
Qy	186	DKKTAIAEYTSRAGEN--GEI--SOLDVDGKEIINEGEVFNFSYLLKKVTI-----PTGYK	236
Db	152	DDREK-AEFLKSAKSPAGIIGNQIRTDQKFM---GVFDES LKERQEA EKNGSGPTGCD	206
Qy	237	HIGQDAFV----DNKNIAEVLPSLETISDYAFHLAKQIDLPDNLKALGELAFFNQ	292
Db	207	WL--DIFLSFTFDKQSSDKV-----EAINOEPLPHVQ-----PD-----IATSTTH	246
Qy	293	ITGKLSLPOLMRLAERA-----FKSNHIKTTIEPRG-----NSLKVIGEASFOQND	338
Db	247	IOGLPPESRDL--JDERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLI-----HNNT	298
Qy	339	LSQLML--PDGLEKIE--SEAFNGNP-----DDHYNNRVVLWTSGKGNPSG	381
Db	299	LSSVLMGSHDGIPEKVSLLYAGNGGFGCAKHDNATVGYKQOQGNVATIINVMKNGSG	358
Qy	382	LAT---ENTVYNPKLSWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRNK--NL	436
Db	359	LVIAGGEGKINP-----SFVLYKEDQLTSGRALSQSEIQNKIDFM	400
Qy	437	EIPKOHGV--TITEIGDNAPRN--VDPONKTLRYDLEEVKLPSTIRKIG--AFAPQSN	490
Db	401	EFLAQNNAKLDLSSEKEKEPRNEIKDFQ-KDSKPY-----LDALGNDRIFAYSK	449
Qy	491	NLKSFEASDDEEIKEG-----AFMNNRIETLELKLVTIGDAAPHINHIYAIVLPE	543
Db	450	--KDPKHSALITEFNKGDLSTLKVMGKKQIKALD-REKNVTLOQNLKH-----D	496
Qy	544	SVEIGRSAPRQNGANNLIFMGSKVKITLGEMAFLSNRLEHLDLSBQK-----Q	591
Db	497	GVMFVNTSNFYKTASK-----SPNKGVG-----VTNGVSHLEAGFSKVAVFNLNLNLA	547
Qy	592	LTEIPVQAFSDNALKEVLLPASLKTIRBEAPKONHLKOLEVASALSHTAFNALDNDGDE	651
Db	548	ITSVVRDLEDKLIAKGLSPQEA NKLVKDFLSSN--KEL--VGKALNF-----	591
Qy	652	QFDNKVVVTHNSYALADGHHFIVDPDKLSTIVDLEKIILKLEGLDYSLRTQTTQTF	711
Db	592	---NKAVAEAKNTG-----NYDEVRAQOKLEKSLKQEHLE-----	625
Qy	712	RDMTTAGKALLSKNLROGEKQKFLQEAQFFLGRVLDLKA TAK EK-----ALVTKKATKN	767
Db	626	KDV---AKNLESKS-----GNKNQWEAKAQ-----ANSQDIFAILINKENARD	666
Qy	768	GQLLE-----RSINKAVLAYNNSAI-----KKANVKELEKELDLLTLGV	806
Db	667	ARATAYAQLNGIKREISDKLENINKOLKDFSKSFDGFGKGNKDFSKAEETLKALKGSV	726
Qy	807	EGKGPLAQATWQGVYLLKTP---LPLPEYVYIGLVNVPDKSGKLIYALDMSDTTIGEGQD	863

727	Db	KDLG-----INPWIISKVENLALNEFKNGKVKDFSKVTOA-----KSD-----QEN	769
864	Qy	AVNGPILN-----VDENEGYHALAVATLA-DVEGL-----DIKTILNKLQSLSIRQ	911
770	Db	SIKDVIINQIKTIDKVDLNLQ-----AVSVAKIACDFSGVEGALADLKNFSKBLAQ--QAQK	824
912	Qy	VPTAAAYHRAGIFQAION	928
825	Db	NESEFNNGKSEIYOSVKQ	841

## RESIII.T 44

```

PCT-US93-09782-2
Sequence 2, Application PC/TUS9309782
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: TUMURO, MURALI K.R.
TITLE OF INVENTION: THE taga GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09782
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-09782-2

```

Query Match	3.3%;	Score 167.5;	DB 5;	Length 859;
Best Local Similarity	20.4%;	Pred. No. 0.0032;		
Matches 212;	Conservative 150;	Mismatches 344;	Indels 331;	Gaps 59;

Qy	27	LVKEPILKQTOASSISGADYABSGSKL	KINETS	GPVDDTV	TDLS	DRKTTPEK	IKON	86
Db	1	MTNETIDQQPQTEAAAFNQFQFIN	LQVAFK	----	VDNAV	SDPDQK	PVDK	49
Qy	87	LAGPREQELKAVTENTESEKITSGS	OLEQES	KLNTV	----	PSTSNWEI	COFITKGN	144
Db	50	-----NDRDNFQAFEG--ISQ	REYS	----	NKAIKNPT	KQYFSD	PINKSN	91
Qy	145	TLVG-----LSKSGVEKLSQTDH	LVLP	QAADGT	QLIQV	ASF	----	TP
Db	92	DLINKNLIVBSSTKSFQKDGQRY	RIFTS	SWSHQND	PSKIN	TRCIRNF	MEHTIO	151
Qy	186	DKKTAIAEYTSRAGEN--GEI--	SOLD	VDGKEI	NEGEVENS	YLLK	KVTI	236
Db	152	DOKEK--AEPLKSAKOSFAGII	IGNQIR	TDQFM	----	GVFDES	LKEREAK	206
Qy	237	HIGQDAFV-----DNKNIAEVL	NP	SL	ETISD	YAF	LAHLAK	292



207 WL--DIFLSFTFDKQSDVK-----EAINQEPHVVQ-----PD-----TATSTH 246  
293 ITGKLSLPRQLMRLAERA-----FKSNHKTIEFRG-----NSLKVIGEASFQND 338  
247 IQGLPPESRDL--LDERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLI-----HNNT 298  
339 LSQMLM--PDGLEKIE--SEAFGTGNGP-----DDHYNRVVLWTKSGKPNPSG 381  
299 LSSVLMGSHDIEPEKVSLLYAGNGGPGAGXHDNATVGYKQQGNNVATTINVMKNGSG 358  
382 LAT---ENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLQVKRKN--NL 436  
359 LVITAGGEKGINNP-----SFYDKEDQITGSORALSQBEIQNKIDFM 400  
437 EIPKQHNGV--TITEIGNAFRN--VDFQNKTLRKYLEEVKLPSTIRKIG--AFAPQSN 490  
401 EFLAQNNAKLDSLSKEKEFRNEIKDFQ--KDSKPY-----LDALGNDRIFA VSK 449  
491 NLKSPESDDELEIEKEG-----AFMNRITETLEKDKLVITIGDAAPHINHIIVLPE 543  
450 --KDPKHSALITFENKGDLSYTLKVMGKKQIKALD--REKNVTLOGNLKH-----D 496  
544 SVOBGRSAFRQNGANNLIFMGSKVKTIGEMAFNSRLEHLDLSEQK-----Q 591  
497 GVMFVNVSNFKYTNASK-----SPNKGVG-----VTNGVSHLEAGFSKVAVPNLNLNLA 547  
592 LTEIPVQAFSDNALKEVLPASLKTIRBEAFKKNHLKQLEVASALSHTAFNALDDNDGDE 651  
548 ITSVVRDLEDKLIAGKLSPOEANKLVKDFLSSN--KEL--VGKALNF-----591  
652 QFDNKVVVTHNSVALADGEHFIVDPDKLASTIVDLSEKILKLEGLDYSLRTQTQTF 711  
592 ---NKAVAEAKNTG-----NYDEVKRAQKDLSEKSLKGRHLE-----625  
712 RDMTAGKALLSKNLROGEKQFLOEQAQFPLGRVLDLQAKAEK-----ALVTKKATKN 767  
626 KDV--AKNLESKS-----GNKNMEAKAQ-----ANSQIBFALINKEANRD 666  
768 QQLAE-----RSINKAVLAYNNSAI-----KCANVKELEKELDLITGLV 806  
667 ARALAYAQNLKGIKRELSKLENINKDLKDSKSPDFGKNGKNDPFSKAEETLKALXGSV 726  
807 EGKQPLAQATWQGVYLLKTP---LPLPEYIGLVYFPDKGKLIYALDMSDTTIGEGQK 863  
727 KDLG-----INPEWISKVENLAALNEFKNGKNDPFSKVTQA-----KSD-----QEN 769  
864 AVGRPILN-----VDEDNEGHALAVATLA--DYEGL-----DIKTILNSKLSQTSIRQ 911  
770 SIKVILNQKITDKVDELNQ---AVSVAKIACDFSGVEQALADLKNFSEQLAQ--QAOK 824  
912 VPTAAYHRAGIFQATQN 928  
825 NESFNVGKSEIYQSVCN 841

## RESULT 45

US-08-053-614-4

; Sequence 4, Application US/08053614

; Patent No. 5403924

; GENERAL INFORMATION:

; APPLICANT: COVER, TIMOTHY L.

; APPLICANT: BLASER, MARTIN J.

; APPLICANT: TUMURU, MURALI K.R.

; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING

; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NEEDLE &amp; ROSENBERG, P.C.

; STREET: 133 Carnegie Way, Suite 400

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: USA

ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/053,614  
FILING DATE: 19930426  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SPRATT, GENDOLYN D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 2200.009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1181 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-053-614-4

Query Match 3.3%; Score 167.5; DB 1; Length 1181;

Best Local Similarity 20.4%; Pred. No. 0.0052;

Matches 212; Conservative 150; Mismatches 344; Indels 331; Gaps 59;

Qy 27 LVKEPILKQTCASSISGADYAESGSKLKNETSQVDDTVTDLSDKRTTEPKIKDN 86  
Db 1 MTNETIDQQPTEAFAFPQPFNNLQVAFK-----VDNAVASYDDQKPIYDK---- 49  
Qy 87 LAKGPREQKALVTEENTSEKQITSGSQLESKESLSNKTV--PSTSNWEICDFITKGN 144  
Db 50 -----NDRDNRAQEG--ISQLEREYS--NKAIKPTKKNQYFSPFINKSN 91  
Qy 145 TLVG-----LSKGVKLSQTDHLVLPQAADGTQLIQVASFAP-----TP 185  
Db 92 DLINKDNLI VVESSTKSFQFGDQRYRIFTSWSHQNDPSKINRCIRNPMHTIQPIP 151  
Qy 186 DKTAIAEYTSRAGEN--GEI--SOLDVDGKEIINEGEVFNYSLLKVTI-----PTGYK 236  
Db 152 DDKEK--AEFLKSAQSPAGIIGNQIRTDQFM---GVFDESLEKQEAKEKNGPTGGD 206  
Qy 237 HIGQDAFV---DNKNIAEVNLPESLETISDYAFALHALKQIDLPDLNKAIGELAFDQ 292  
Db 207 WL--DIFLSFTFDKQSDVK-----EAINQEPHVVQ-----PD-----TATSTH 246  
Qy 293 ITGKLSLPRQLMRLAERA-----FKSNHKTIEFRG-----NSLKVIGEASFQND 338  
Db 247 IQGLPPESRDL--LDERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLI-----HNNT 298  
Qy 339 LSQMLM--PDGLEKIE--SEAFGTGNGP-----DDHYNRVVLWTKSGKPNPSG 381  
Db 299 LSSVLMGSHDIEPEKVSLLYAGNGGPGAGXHDNATVGYKQQGNNVATTINVMKNGSG 358  
Qy 382 LAT---ENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLQVKRKN--NL 436  
Db 359 LVITAGGEKGINNP-----SFYDKEDQITGSORALSQBEIQNKIDFM 400  
Qy 437 EIPKQHNGV--TITEIGNAFRN--VDFQNKTLRKYLEEVKLPSTIRKIG--AFAPQSN 490  
Db 401 EFLAQNNAKLDSLSKEKEFRNEIKDFQ--KDSKPY-----LDALGNDRIFA VSK 449  
Qy 491 NLKSPESDDELEIEKEG-----AFMNRITETLEKDKLVITIGDAAPHINHIIVLPE 543  
Db 450 --KDPKHSALITFENKGDLSYTLKVMGKKQIKALD--REKNVTLOGNLKH-----D 496  
Qy 544 SVOBGRSAFRQNGANNLIFMGSKVKTIGEMAFNSRLEHLDLSEQK-----Q 591  
Db 497 GVMFVNVSNFKYTNASK-----SPNKGVG-----VTNGVSHLEAGFSKVAVPNLNLNLA 547

```
QY 592 LPEIPVQAFSDNALKEVLLPASLKTIRREAFKKNHLKOLEVASALSHIAFNALDDNDGDE 651
DB 548 ITSVRRDLEDKLIAGLSPOEANKLVKDFLSSN--KEL-VGKALNF-----591
QY 652 QPDNKVVVTHNSYALADGEHFIIVDPDKLSSTIVDLKILKLEGLDYSTLRQTQTQF 711
DB 592 ---NKAVAEAKNTG-----NYDEVKRAQKDLKSLKREHLE-----625
QY 712 RDMTTAGKALLSKNLROGEKQKFLQEAQFFLGRVLDLKAIAKAEK-----ALVTKKATKN 767
DB 626 KDV---AKNLESKS-----GNKNWEAKAQ-----ANSQDEIFALINKENRD 666
QY 768 GQLE-----RSINKAVLAYNNSAI-----KXANYKRLKEKLDLLTGLV 806
DB 667 ARAIYAQNLKGIKRELSKLENINKDLKDFSKSPDFGKNGKNKDFSKAEETLKALKGSV 726
QY 807 EKGGLAQATWQGVYLLKTP---LPLPEYIYGLNVYFDKSGKLIYALDMSDTIGEGQKD 863
DB 727 KDLG-----INPEWISKVENLNAALNEFKNGKNKDFSKVTQA-----KSD-----QEN 769
QY 864 AVGNPILN-----VDEDNVGHALAVATLA-DYEGLE-----DIKTILNSKLSQTSIRQ 911
DB 770 SIKQVILNQKITDKVDELNQ---AVSVAKIACDFSGVEQALADLKNFSKEQLAQ--QAOK 824
QY 912 VPTAAYHRAGIFQAION 928
DB 825 NESFNVGKSEIYQSVKN 841

RESULT 46
US-08-316-397B-4
; Sequence 4, Application US/08316397B
; Patent No. 5733740
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,397B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/588-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1181 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-316-397B-4
```

Query Match

3.3%; Score 167.5; DB 1; Length 1181;

```
Best Local Similarity 20.4%; Pred. No. 0.0052;
Matches 212; Conservative 150; Mismatches 344; Indels 331; Gaps 59;
QY 27 LVKEPILKQTQASSISGADYAESGSKLKLNETSGPVDVTVDLFSDKRTTPEKIDN 86
DB 1 MTNETIDQOPOTEAAPNFQPFINNIQVAFK-----VDNAVASYDPQDKPIVDK----49
QY 87 LAKGPREQELKAVTENTSEKQITSGSOLEQSKESLSLNTKV--PSTSNWEICDFITKGN 144
DB 50 -----NDRDNROAFEG--ISQLREEYS-NKAIKNTFKKQYFSDFLKNSN 91
QY 145 TLVG-----LSKSGVEKLSQTDHLVLPQAAOGTQLOIVASPAF-----TP 185
DB 92 DLINKDNLIIVBESSTKSFQKFGDQRYRIFTSWVSHQNDPSKINTRCIRNFMEHTPIPIP 151
QY 186 DKKTAAEYTSRAGEN--GEI---SOLDVDGKEIINEGEVFNYSLLKVTI-----PTGYK 236
DB 152 DDKEK-ABFLKSAKQSPAGIIGNQIRTDQKFM-----GVFDESLKEROBAERKGGTGD 206
QY 237 HIGQDAFV---DNKNIAEAVNLPESLETISDYAFALHALKQIDLPNLKAIGELAFDQ 292
DB 207 WL--DIFLSFIDFKKQSSDVK-----EAINQEPHPVQ-----PD-----IATSTH 246
QY 293 ITGKLSLPRQLMLAERA---FKSNHIKTIFFRG-----NSLKVIGEASFDQND 338
DB 247 IQGLPPESRDL--LDERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLI-----HNNT 298
QY 339 LSQML--PDGLEKIE-SEAFITGNPG-----DDHYNNRVVLTWTKSGKNPSG 381
DB 299 LSSVLMSGHDGIEPEKVSLLYAGNGGFGCAKHDWNAVGYKQOQGNNAVATIINVHMKNVG 358
QY 382 LAT---ENTYVNPDKSLMQESPEIDYTKWLEBDFYQKNSVTGFSNGLQKQVKNK--NL 436
DB 359 LVIAGKEGINNP-----SFYLYKEDQLTGSORALSQEIQNKIDPM 400
QY 437 EIPKQHNGV--TITEIGDNAPRN--VDFQNKTLRYDLEEVKLPSTIRKIG--AFAFSN 490
DB 401 EPLAQNNAKLDSLSEKEKEFRNEIKDFQ-KDSKPY-----LDALGNDRIAFVSK 449
QY 491 NLKSEASDDLEEIEKVG-----AFMNRRTLELKKLVITIGDAAPHINHIYAILVPE 543
DB 450 --KDPKHSALITEFNKNGDLSYTLKVMGKKQIKALD-REKNVTLOGNLKH-----D 496
QY 544 SVQETGRSAFRQNGANNLIFMGSVKVTILGEMAFNLRLHLDLSEQK-----Q 591
DB 497 GVMFVNYNFKYTNASK-----SPNKGVG-----VTNGVSHLEAGFSKVAVFNLPLNLA 547
QY 592 LPEIPVQAFSDNALKEVLLPASLKTIRREAFKKNHLKOLEVASALSHIAFNALDDNDGDE 651
DB 548 ITSVRRDLEDKLIAGLSPOEANKLVKDFLSSN--KEL-VGKALNF-----591
QY 652 QPDNKVVVTHNSYALADGEHFIIVDPDKLSSTIVDLKILKLEGLDYSTLRQTQTQF 711
DB 592 ---NKAVAEAKNTG-----NYDEVKRAQKDLKSLKREHLE-----625
QY 712 RDMTTAGKALLSKNLROGEKQKFLQEAQFFLGRVLDLKAIAKAEK-----ALVTKKATKN 767
DB 626 KDV---AKNLESKS-----GNKNWEAKAQ-----ANSQDEIFALINKENRD 666
QY 768 GQLE-----RSINKAVLAYNNSAI-----KXANYKRLKEKLDLLTGLV 806
DB 667 ARAIYAQNLKGIKRELSKLENINKDLKDFSKSPDFGKNGKNKDFSKAEETLKALKGSV 726
QY 807 EKGGLAQATWQGVYLLKTP---LPLPEYIYGLNVYFDKSGKLIYALDMSDTIGEGQKD 863
DB 727 KDLG-----INPEWISKVENLNAALNEFKNGKNKDFSKVTQA-----KSD-----QEN 769
QY 864 AVGNPILN-----VDEDNVGHALAVATLA-DYEGLE-----DIKTILNSKLSQTSIRQ 911
DB 770 SIKQVILNQKITDKVDELNQ---AVSVAKIACDFSGVEQALADLKNFSKEQLAQ--QAOK 824
QY 912 VPTAAYHRAGIFQAION 928
DB 825 NESFNVGKSEIYQSVKN 841
```

```

Db      825 NESFVNGKSEIYQSVKN 841

RESULT 47
US-09-034-306-4
; Sequence 4, Application US/09034306
; Patent No. 5876943
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,306
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,397
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1181 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-034-306-4

Query Match      3.3%; Score 167.5; DB 2; Length 1181;
Best Local Similarity 20.4%; Pred. No. 0.0052;
Matches 212; Conservative 150; Mismatches 344; Indels 331; Gaps 59;

Qy      27 LVKPEILQTOASSISGADYAESGKGLKINETSGPVDVTDLPFSDKRTTPKIKDN 86
Db      1 MTNETIDQPTQTEAFNQQFINNLOVAFLEK-----VDNAVASYDPQKPIVDK---- 49
Qy      87 LAKGPREQELKAVTENTSEKQITSGSQLESKESLSLNTKV--PSTSNWEICDPITKGN 144
Db      50 -----NDRDNRQAPEG--ISQUREEYS-NKAIKNTKKNQVFSDFINKSN 91
Qy      145 TLVG-----LSKSGVEKLSQTHLVLPQAAAGTQLIQVAF-----TP 185
Db      92 DLINKNDLIVVESSTKSPKQFGDQRYRIFTSVSHQNDPFSKINTRCIRNFMEHTIQPIP 151
Qy      186 DKTKTAIEVTSRAGEN--GEI--SOLDVDGKEIINEGEVFNFSYLLKVKTI-----PTGVK 236
Db      152 DDKEK-AEFLKSAKQSPAGIIGNIRTDQKFM-----GVFDESLEKEREAKNGGPTGCD 206
Qy      237 HIGQDAFV-----DNKNIAFVNLPESLETISDYFAHLAKQIDLPDNLKAI GELAFPNQ 292
Db      207 WL--DIFLSFIPDKQSSDVK-----EAINQEPPLPHVQ-----PD-----IATSTTH 246

293 ITGKLSLPROLMRLAERA-----FKSNHIKLTIEPRG-----NSLKVTGEASPDND 338
247 IQGLPPPSRDL--LDERGNFSKFTGLDMEMLDVEGVADMDDPNYKFNQLLI-----HNNT 298
339 LSQMLM--PDGLEKIE--SEAFITGNPG-----DDHYNNRVVLWTKSGKNPSG 381
299 LSSVLMSGSHDGEIEPEKVSLLYAGNGGFGCAKHDNATVGYKQOQGNVATIIINVEHMGSG 358
382 LAT---ENTYVNPDKSLWQSPBIDYTKWLEEDFTYQKNSVTGSNKLQKVKKNK--NL 436
359 LVIAGEGKGINNP-----SFYLYKEDQLTGSQALSQBEIQNKIDFM 400
437 EIPKOHGV--TITEIGNAFRN--VDFONKTLRKYDLEEVKLPSTIRKIG--AFATQSN 490
401 EFLAQNAKLDSLSEKSEKFRNEIKDFQ--KDSKPY-----LDALGNDRIAPVSK 449
491 NLKSFASDDLEIEKEG-----AFMNNRIETLELKDCLVTIGDAAPHINHIYAIVLPE 543
450 --KDPKHSALITEFNGKGLSYTLKVMGKKQIKALD-REKNVTLOGNLKH-----D 496
544 SVQEIGRSAPRONGANNLIFMGSKVTLGEMAFNSRLEHLDLSEQ-----Q 591
497 GVMFVNTSNPKYTNASK-----SPNKGVG---VTNGVSHLEAGFSKVAVFNLNLA 547
592 LTEIPVQAFSDNALKEVLLPASLKTIRREAFKKNHLKOLEVASALSHTAFNALDDNDGDE 651
548 ITSVVRRDLEDKLIAGLSPOEANKLVKDFLSSN--KEL--VGKALNF----- 591
652 QFDNKVVVKTHTNSYALADGHEHFTVDPDKLSSTTVDELEKILKLEGLDYSTRQTOTQF 711
592 ---NKAVAEAKVTG-----NYDEVKBAQKLEKSLKKEHLE----- 625
712 RDMTTAGKALLSKNLRQGEKQKFLQBAQFPLGRVDLDKATAKAEK-----ALVTYKATKN 767
626 KDV---AKNLESKS-----GNKNKMEAKQ-----ANSQKOEIPALINKEANRD 666
768 GQLE-----RSINKAVLANNNSAI-----KXANVRKLEKELDLTLGLV 806
667 ARATAYAQNLKGIKRELSDKLENINKOLKDFSKFDGFKNGKNKDFSKAEETLKALKGSV 726
807 EGKGLAQATWVQGVYLLKTP---LPLPEYVIGLVNFDKSGKLIYALDMSDTTIGEGQKD 863
727 KDLG-----INPEWISKVENLNAALNEFKNGKKNKDFSKVTQA-----KSD-----QEN 769
864 AYGNDPILN-----VDENEGYHALAVATLA-DYEGLE-----DIKTILNSKLSQTSIRQ 911
770 SIKDVIINQKITDKVDLNLQ---AVSVAKIACDFSGVEQALADLKNFSKEQLAQ--QAQK 824
912 VPTAAVHRAGIFOAIQN 928
825 NESFVNGKSEIYQSVKN 841

RESULT 48
US-09-259-437-4
; Sequence 4, Application US/09259437
; Patent No. 6153390
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/259,437  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,397  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: SPRATT, GWENDOLYN D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 2200.030  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1181 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-259-437-4

Query Match 3.3%; Score 167.5; DB 3; Length 1181;  
Best Local Similarity 20.4%; Pred. No. 0.0052;  
Matches 212; Conservative 150; Mismatches 344; Indels 331; Gaps 59;

QY 27 LVKPEILKQTOASSISGADYAESGSKLKNINETSQPVDDTVTDLFSKRTTPEKIKON 86  
DB 1 MTNETIDQOPQTEAFNPQFINNLQVAFLK-----VDNAVASYPDPQKPIVDK---- 49  
QY 87 LAKPREQELKAVTESEKQITSGSQLEBSKESLSLNTKV--PQTSNWEICDFITKGN 144  
DB 50 -----NDRNROAFEG--ISQLEEYS-NKAIKNPTKQNYFSDFINKSN 91  
QY 145 TLVG-----LSKSGVEKLSQTDHLVLPQQAADGTQLIOVASPAF-----TP 185  
DB 92 DLINKDNLIVSESTSKSFQKFGDQRTFTSWSHQNDPFSKINTRCIRNMEHTIOPPIP 151  
QY 186 DKRTAJAAYTSRAGEN--GEI--SQLDVQKEIINEGEVFNYSLLKKVTI-----PTGYK 236  
DB 152 DDEK-AEFLSAKQSFAGIIGNQIRTDQKFM---GVFDESLEKQEAERKNGGPTGSD 206  
QY 237 HIGQDAFV---DNKIAEVLNPLESETISDYAFALHAKQIDLPNLKAIAGELAFPDQ 292  
DB 207 WL--DIFLSFIDFKQSSDVK-----EAINQEPLPHVQ-----PD-----IATSTTH 246  
QY 293 ITGKLSLPQMLAERA---FKSNHIKTIIEPRG-----NSLKVIGEASFQDND 338  
DB 247 IQLPPESDL--LDERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLI-----HNNT 298  
QY 339 LSQJML--PDGLEKIE--SEAFNGP-----DHNINRVVLWTKSGKNPSG 381  
DB 299 LSSVLGSHDGIPEKVSLLYAGNGGFGAKHDWNAVGYKQDQGNVATIIVHVMGSG 358  
QY 382 LAT---ENTVNPDKSLWQESPEIDYTKMLEDFTYQKNSVTGFSNGLOKVRNK--NL 436  
DB 359 LVIAEGKEGINNP-----SPFLYKEDQLTGSQRLSQEETQNKIDFM 400  
QY 437 EIPKQNGV--TITEIDNAFRN--VDFQNTLRYKDYDLEVKLPSTIRKIG--AFASFQSN 490  
DB 401 EFLAQNNAKLDLSSEKEKEFRNEIKDFQ--KOSKPY-----LDALGNDRIAFVSK 449  
QY 491 NLKFSFASDDELEIKEG-----AFWNNRIETLELKDVLVTIGDAAFHINHIYAVLPE 543  
DB 450 --KDPKHSALITFENKGDLSYTLKVMGKKQIKALD--REKNVTLOGLNKH-----D 496  
QY 544 SVOEIGRSAPRONGANNLIFMGSKVITLGEAFNLNRLEHLDLSEOK-----Q 591  
DB 497 GVMFVNYSNFKYTNASK-----SPNKGVG-----VTNGVSHLEAGFSKAVFNLNLLNLA 547

QY 592 LTEIPVQAFSDNALKEVLLPASLKTIRBEAFKKNHLKOLEVASALSHIAFNALDNDGDGE 651  
DB 548 ITSUVRRDLEDKLIAKGLSPQEAANKLVKDFLSSN--KEL-VGKALNF-----591  
QY 652 QPDNKVVVTHNSVALADGEHFIVDPKLSSTIVDLEKILKILEGLDYSTLRQTTQTOF 711  
DB 592 ---NRVAEAKNTG-----NYDEVKRAQDKLEKSLKREHLE-----625  
QY 712 RDMTTAGKALLSKSNLROGEKOFIQAQFFLGRVDLDKAIKAEK---ALVTKKATKN 767  
DB 626 KDV-----AKNLESK---GNKNMEAKAQ-----ANSQDEIFALINKEND 666  
QY 768 GOLLE-----RSINKAVLAYNNSAI-----KCANVYKRLKELDLTLGLV 806  
DB 667 ARAIAVAQNLIKGIKRELSKLENINKDLKDFSKPDGPKNGKNKDFSKAEETLKALKGSV 726  
QY 807 EKGGLPQAQTVQGVYLLKTP---LPLPEYIIGLVYFDKSGKLIYALDMSDTIGEGQKD 863  
DB 727 KDLG-----INPEWISKVENLNAALNEFKNGKNKDFSKVTQA-----QEN 769  
QY 864 AVGNPILN-----VDEDEGVAHALAVATLA-DYEGE-----DIKTILNSKLSQTSIRQ 911  
DB 770 SIKDVIINQKITDKVDELNQ---AVSVAKIACDFSGVEQALADLNKFSKBQLAQ--QAK 824  
QY 912 VPTAAVHRAGIFQAION 928  
DB 825 NESFNVGKSEIXQSVKN 841

RESULT 49  
PCT-US93-09782-4  
Sequence 4. Application PC/TUS9309782  
GENERAL INFORMATION:  
APPLICANT: COVER, TIMOTHY L.  
APPLICANT: BLASER, MARTIN J.  
APPLICANT: TUMURU, MURALI K.R.  
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING  
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
STREET: 133 Carnegie Way, Suite 400  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/09782  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: SPRATT, GWENDOLYN D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 2200.009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1181 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-09782-4  
Query Match 3.3%; Score 167.5; DB 5; Length 1181;  
Best Local Similarity 20.4%; Pred. No. 0.0052;



Db 136 YQNLVFNMDYQDLNLSGCHRRHRETDYDT- - - - - EKWFESH- - - - - 174  
Qy 171 DGTQLIQVASFAPTPDKKTAIAEYTSRAGENGESQLDVGKELINEGE--VFN5YLLKK 228  
Db 175 --DQKNVSIYA--NQTSYCWMKDYFNKNVNNYDHLNVSINRLEAEAFYAFDDFSQTI 229  
Qy 229 VTPTGYKHIGODAFVDNK- - - - - NIAEV- - - - - NLPESLET 260  
Db 230 KLTNNSYQTVNIDVFNFDNLCILALRFLSLERFNLNRSYTRNQYNFEKIGELLET 289  
Qy 261 ISDYAFALHAKQIDLPDNLKAGELAFPDNQITGKLS- - - - - 298  
Db 290 IFAVFVSHRHLQGIHLQVPCFAFYLVNSSQISVKDSQIQVYSFSDTLKLVDTNKVQDY 349  
Qy 299 - - - - - LPROLMRLAERAPKSNHKTIEPRGNSIKVIGEAS- - - - - PQNDL 339  
Db 350 FKFLQEPFR-LTHVSQAIPVSNATNAVENLVLLKVKVGHANLVNIPQTFNDFDFVNL 408  
Qy 340 SOLMLPDGLE- - - - - KIBSEAFPGDGDHYNRNVLVWTKSGKNPSGLATE-NTYVNP 391  
Db 409 QHLKLEFLEPNILTKQLEN- - - - - LLSIKOSKNLKFRLNFYTYVAQ 453  
Qy 392 DKSLWQSPIDYTKLMBEDFTQKNSVTGFSKGLQKVRNKNLEIPKOHNGVTTTEIG 451  
Db 454 ETSRKO- - - - - ILKOATT- - - - - IKNLKNKNQEBETPE- - - - - TKDET 487  
Qy 452 DNAPRVNDFONKTRKYDLEE--VKLPST-- - - - - - IRKIGAPAFQSNLKSFEAS--DDEEI 504  
Db 488 SESTSGKFFDHLSELTEDSDFNVNLOATQIEIVDSLHLKLLIRSTNLKPKLSYKEMEKS 547  
Qy 505 KEGAF--MNNRIETL--ELKDKLVITIGDAAAPHINHIYAVLPESV- - - - - QEI--GR 550  
Db 548 KMDTFIDLKNIYETLNNLKCSVNI--HGNISVELTNKDSFTFYKFLTLNQLQHAH 605  
Qy 551 SAFRQCANNLIPGSKVKTGEMAFLSNRLEHLDSEQQLTEIPVQAFSDNALKEVLL 610  
Db 606 YTFKQNE- - - - - FQFNVK- - - - - SAKIESSSLESLEDIDSLCKSIASCNQLQVNI 652  
Qy 611 PASL- - - - - KTRIEAFKQNLH- - - - - KOLEVASALSHIAFNALDNDGDGEQFNDKVVVKTNN 664  
Db 653 IASLLYPNNIQKPNFKPNLLFFKQFQLKNLVNSINCILD- - - - - 694  
Qy 665 SYALADGEHFIVDPDKLSSTIVLEKILKLEGLDYSTLROTTO- - - - - TQFRD 713  
Db 695 QHILNISELEKXKIKAFI--LKRYLLOYLDYTKLFTLQOQPELNOVINOOLEE 752  
Qy 714 MTAGKALLSKSNLQCEKQK- - - - - FLOBAQFFLGRVLDLKAIA- - - - - KAEKALVTK 762  
Db 753 LTVSE- - - - - VHQVWENHKQKAFYELCEPIKESSQTLQIDFDQNTVSDSFKKILESI 808  
Qy 763 KATNGQOLLERSINKAVLANNNSAIKKNVRLEKELDILTLGLVEGKGLAQTMTVQGY 822  
Db 809 SESKHYHRLNPSQ- - - - - SSSLISEN- - - - - EBIQELLKACDEKGLVKA- - - - - 851  
Qy 823 LKATPLPEYIYGLNVYFD 842  
Db 852 YKFLCLP- - - - - TGYTD 866

## RESULT 51

US-08-851-843A-54

; Sequence 54, Application US/08851843A

; Patent No. 6093809

## ; GENERAL INFORMATION:

; APPLICANT: Cech, Thomas R.

; APPLICANT: Lingner, Joachim

; APPLICANT: Nakamura, Toru

; APPLICANT: Chapman, Karen B.

; APPLICANT: Morin, Gregg B.

; APPLICANT: Harley, Calvin

; APPLICANT: Andrews, William H.

; TITLE OF INVENTION: No. 6093809el Telomerase

; NUMBER OF SEQUENCES: 225  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/851,843A  
; FILING DATE: 06-MAY-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/846,017  
; FILING DATE: 25-APR-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/844,419  
; FILING DATE: 18-APR-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/724,643  
; FILING DATE: 01-OCT-1996  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Apple, Randolph T.  
; REGISTRATION NUMBER: 36,429  
; REFERENCE/DOCKET NUMBER: 015389-002930US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 54:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 872 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
; US-08-851-843A-54

Query Match 3.3%; Score 167; DB 3; Length 872;

Best Local Similarity 20.5%; Pred. No. 0.0036;

Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;

Qy 13 TTVSVVTHNQVFP- - - - - SLVKEPILK- - - - - QTOASSISGADYAESGSKLK 57  
Db 17 TNLDFVLQNLVYKSGIEHYKTOQQQIKEDLKLKFNQDQDGNDDDEENNNSN-K 75  
Qy 58 INETSGPVDDTVDLFSDKRTTPEIKD- - - - - NLAKGPREQBELKAVTENTESEKQIT 110  
Db 76 QQELRRVNIQKQVQLIKKVGSKVEKDLNLEDENKKNGLSEQQVKEBQLATITEQVK 135  
Qy 111 SCSQLEQSKESLSLNKTVPTSNWEICDPTKGNLTGLSKSGVKEKLSQTHLVLPSQAA 170  
Db 136 YQNLVFNMDYQDLNLSGCHRRHRETDYDT- - - - - EKWFESH- - - - - 174  
Qy 171 DGTQLIQVASFAPTPDKKTAIAEYTSRAGENGESQLDVGKELINEGE--VFN5YLLKK 228  
Db 175 --DQKNVSIYA--NQTSYCWMKDYFNKNVNNYDHLNVSINRLEAEAFYAFDDFSQTI 229  
Qy 229 VTPTGYKHIGODAFVDNK- - - - - NIAEV- - - - - NLPESLET 260  
Db 230 KLTNNSYQTVNIDVFNFDNLCILALRFLSLERFNLNRSYTRNQYNFEKIGELLET 289  
Qy 261 ISDYAFALHAKQIDLPDNLKAGELAFPDNQITGKLS- - - - - 298  
Db 290 IFAVFVSHRHLQGIHLQVPCFAFYLVNSSQISVKDSQIQVYSFSDTLKLVDTNKVQDY 349



QY 299 -----LPROLMRLAERAFKSNHKTIEPRGNSLKVIGEAS-----FQDNDL 339  
Db 350 FXFLOEPPR-LTHVSOQALPVSATNAVENLVLLKKVKHANLVLSIPTQFNDFYFVNL 408  
QY 340 SOLMLPDGLE-----KIESBAFTGNPGDDHYNRVLWTKSGKNPGLATE-NTYVNP 391  
Db 409 QHLKLEFGLPILTKQLEN-----LLLSIKQSNLKLRLNFYTYAQ 453  
QY 392 DKSLWOESPEIDYTKWLEBEDFYQNSVTFGSKGLQKVRNKNLEIPKQHGNGVITEIG 451  
Db 454 ETSRKQ-----ILQOATT-----IKLNKNKQOETPE-----TKDETP 487  
QY 452 DNAFRNVDFQNTKLARKYLEE--VKLPST---IRKIGAFAFOSNLLKSPAS--DDLEI 504  
Db 488 SESTSGMFPFDHLSBELTELEDFSVNLQATQBYDLSLKLIRSTNLKPKFKLSYKYMESKS 547  
QY 505 KEGAF--MNNRIETI--ELKDKLVITGDAAPHINHIYAVLPESV-----QBI--GR 550  
Db 548 KMDTFIDLKNIYETLNNLKRCVNI--HGNISYELTNKDSYFKFKLTNLQELQAHK 605  
QY 551 SAFRONGANNLIPMGSKVTLGEMAPLNRHLDLSEOKLQTEIPVQAFSDNALKEVLL 610  
Db 606 YTFKQNE-----FQFNVK-----SAKIESSLESLEIDSLCKSTASCKNLQVNI 652  
QY 611 PASL---KTIREAFKQHL---KQLEVASALSHIAFNALDNDGDEQFDNKKVVVYKTHN 664  
Db 653 IASLLYPNNIQNPNKPNLLFPKQFQKLNENVSINCILD-----TQPRD 713  
QY 665 SYALADGHEFIVDPKLASTIVDLKILKIEGLDYSTLRQTTQ-----TQPRD 713  
Db 695 QHILNSISEFLEKKNKIKAFI--LKRYLLQYLYDTYKLFKTLQQLPELNQVYINQOLEE 752  
QY 714 MTTAKALLSKNLKQGEKQ-----FLOAQFFLRVLDLKAIA---KAEKALVTK 762  
Db 753 LTVSE-----VHKQVWENHKQAFYBPLCEFIKESQTLQIDFQNTVSDSICKILSSI 808  
QY 763 KATKNGQLLERSINKAVLAYNNSAIKKANVKELEKELDLTGLVKGKPLAQATWVQGVY 822  
Db 809 SESKHYHURLPSPQ-----SSSLIKSEN-----BEIQELLKACDEKGLVKA-----851  
QY 823 LKTLPLPEYVIGLNVYFD 842  
Db 852 YKFPFLCLP-----TGTYD 866

## RESULT 52

US-08-974-549A-221  
Sequence 221, Application US/08974549A  
Patent No. 6166178

## GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.  
APPLICANT: Lingner, Joachim  
APPLICANT: Nakamura, Toru  
APPLICANT: Chapman, Katen B.  
APPLICANT: Morin, Gregg B.  
APPLICANT: Harley, Calvin B.  
APPLICANT: Andrews, William H.  
TITLE OF INVENTION: Human Telomerase Catalytic Subunit  
NUMBER OF SEQUENCES: 727  
CORRESPONDENCE ADDRESS:  
ADDRESS: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,549A

FILING DATE: 19-NOV-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/854,050  
FILING DATE: 09-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/911,312  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/912,951  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/915,503  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/17618  
FILING DATE: 01-OCT-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/17885  
FILING DATE: 01-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph Ted  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002610US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 221:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 872 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-974-549A-221

Query Match 3.3%; Score 167; DB 3; Length 872;  
Best Local Similarity 20.5%; Pred No. 0.0036;  
Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;

QY 13 TTVSVVTHNQEVF-----SLVKEPIK-----QTQASSISIGADYAESGSKLK 57  
Db 17 TNLDFVLQNLVYKQSIHYKTKQQQIKEDLKLKFKNQDQDGNNGDDDEENNSN-K 75  
QY 58 INETSGPVDDTVTLFSDKRTTPKIKD-----NLAKGPBEQLKAVTENTSEKQIT 110  
Db 76 QOELLRRVNIQKQVQLIKYKGVSKVEKDLNLDENKNGLSRQGVKEEQRLTITEQVK 135  
QY 111 SGSOLEQSKESLSLNTKVPSTSNWEICDFITKNGTLVLGSKSGVEKLSQTLHLVLPQAA 170  
Db 136 YQNLVFNMDYQDLNLSGCHRRHRETDYDT-----EKWFEISH-----174  
QY 171 DGTQLIQVAPFATPKDKTAIAYTSRAGENGEISQLDVGCKEINEGE--VFNSYLLKK 228  
Db 175 --DQKNYVSIYA---NQKTSYCWMLKDYFNKNYDHLNVSNRLTEAEFYAFDDFSQTI 229  
QY 229 VTITGYKHICQDAFVKNK-----NTAEV-----NLPLESET 260  
Db 230 KLTNNSYQTVNIDVNFNLCILALLRPLLSLRFNLIRSSYTRNOYNPEKIGELLE 289  
QY 261 ISDYAFALHAKQIDLPDLNKAIGELAFDFNQITGKLS-----298

```
Db 290 IPAVFVSHRHLOGIHLQVPCFAFOYLVNSSSQISVKDSQLOVVSFSTDCLKLVDTNKVQDY 349
Qy 299 -----LPRQLMLAERAFKSNHIKTIIEFRGNSLKVIGEAS-----FQDNDL 339
Db 350 KFXLQEPFR-LTHVSQQAIPVSATNAVENLVLLKKVKHANLVLSIPTQPNFDFYFVNL 408
Qy 340 SOLMLPDGLE-----KIESEAFGNPDGDDHNNRVLVWTKSGKNPSGLATE-NTYVNP 391
Db 409 QHLKLEFGLFNILTKQLEN-----LLLSIKQSKMLKFLRLNFYFYVAQ 453
Qy 392 DKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVKRNKNLEIPKQHNGVTITEIG 451
Db 454 ETSRKO-----ILKOATT-----IKLNKNKQBEETPE-----TKDETP 487
Qy 452 DNAFRNVDFQNTLRKYDLE-----VKLPST-----IRKICAFAFQSNLNKSPAS--DDLREI 504
Db 488 SESTSCMKFFDHLSELTELEDFSVNLQATQEIYDSLHLKLLIRSTNLNKKFLSVKYEKMS 547
Qy 505 KEGAF--MNNRIETL-ELKDKLVITIGDAAFINHIYAVLPSV-----QEI--OR 550
Db 548 KMDTFIDLKNIYETLNNLRKCSVNI SNP--HGNISYELTNKDSFTFYKFKLTNLNQLQHA 605
Qy 551 SAPRONGANNLFWGSKVKTLCGEMAFLSNRLEHLDLSEQOLTEIPVQAFSDNALKEVLJ 610
Db 606 YTFKQNE-----QFNNVK-----SAKIESSLESLEDIDSLCKSIASCNKLQVNI 652
Qy 611 PASL-----KTIREAFKKNHL-----KQLEVASALSHIAFNALDNDGDGEQFDNKKVVKTHN 664
Db 653 IASLLYPNNIKNPFPKNLLPFQEPQOLKNLENSVINCILD-----694
Qy 665 SVALADGEHPIVDPKLSSTIVDLEKILKLEGLDYSTLRQTTQ-----TQPRD 713
Db 695 QHLNISSEFLKGNKKIKAPI--LKRYLLQVYLDYTKLFLKQLQPELVNQQVYNQLEE 752
Qy 714 MTAGKALLSKNLROGEKOK-----PLQEAQFFLCGRVLDKAIK---KAERKALVTK 762
Db 753 LTUSE-----VHKQVWENHKKQAFYEPCEFIKESQTLQIDFDQNTVSDDSIKKILESI 808
Qy 763 KATKNGQLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLVEGKPLAQATMVQGVY 822
Db 809 SESKYHYLRNPSQ-----SSSLIKSEN-----EIQELKACDEKGVLVKA-----851
Qy 823 LKTPPLPEYIGLVNVPD 842
Db 852 YVKFPLCLP-----TGYYD 866

RESULT 53
US-08-854-050-8
; Sequence 8, Application US/08854050
; Patent No. 6261836
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6261836el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,050
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-0029300S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-08-854-050-8
```

```
Query Match 3.3%; Score 167; DB 3; Length 872;
Best Local Similarity 20.5%; Pred. No. 0.0036;
Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;
```

```
Qy 13 TTVSVTHNQEVF-----SLVKEPIK-----QTOASSISGADYAESKSKUK 57
Db 17 TNLDFVLQNLVYKSIQIEHYKTTQOOQIKEEDLKLFKNQDQDQSGNDDDEENSN-K 75
Qy 58 INETSGPVDDTVDLFSDKRTTPEKID-----NLAKPREQELKAVTENTESEKQIT 110
Db 76 QOELLRRVNIQKQVQLIKKVGSKVEKDLNLEDNKNKNGLSEQQVKEQLRTITEQVK 135
Qy 111 SCSQLEQSKESLSLNKTVPSTSNWBEICDFTKGNLTGLVLSKSGVEKLSQTDHLVLPQAA 170
Db 136 YQNLVFNMDYQLDLNESGGHRRHRETDYDT-----EKWFEISH-----174
Qy 171 DGTQLIQVASPAFTPKDTAIAEYTSRAGENGESIQLDVDGKEIINEGE--VFNSYLLKK 228
Db 175 --DQKNYYSIYA--NQKTSYCWMLKDYFNKNYDLNVSINRLETEAEFYAFDQFSQTI 229
Qy 229 VTIPGTGKHIGODAFVDNK-----NIAEV-----NLPSLET 260
Db 230 KLTNNSYQTVNIDVNFNNLCILALLRFLLSLERNILNIRSSYTRNQVNFKEIGELLE 289
Qy 261 ISDYAFALHAKQIDLPNLNKAIGELAFDFNQITGKLS-----298
Db 290 IFAVVFSHRHLOGIHLQVPCFAFOYLVNSSSQISVKDSQLOVVSFSTDCLKLVDTNKVQDY 349
Qy 299 -----LPRQLMLAERAFKSNHIKTIIEFRGNSLKVIGEAS-----FQDNDL 339
Db 350 KFXLQEPFR-LTHVSQQAIPVSATNAVENLVLLKKVKHANLVLSIPTQPNFDFYFVNL 408
Qy 340 SOLMLPDGLE-----KIESEAFGNPDGDDHNNRVLVWTKSGKNPSGLATE-NTYVNP 391
Db 409 QHLKLEFGLFNILTKQLEN-----LLLSIKQSKMLKFLRLNFYFYVAQ 453
Qy 392 DKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVKRNKNLEIPKQHNGVTITEIG 451
```

```
Db 454 ETSRKQ-----ILQATT-----IKLNKNNKQETPE---TKDTP 487
Qy 452 DNAFNVDQNKTLRYDLE--VKLPST---IRKIGAFQSNLKSFEAS--DDLBEI 504
Db 488 SESTGKMFDFHLSLTELDFSVNLQATQEIYDSLHLKLLIRSTNLKPKLSYKYEKMS 547
Qy 505 KEGAF--MNNRIETL-ELKDKLVTTGDAAFHINHIAIVLPESV-----QBI--GR 550
Db 548 KMDTFIDLKNIYETLNNLKRCVSNISNP--HGNISYELTNKDSYFKFKLTINQBLQAH 605
Qy 551 SAFRONGANNLIPMSKVKTLGEMAFNLNLEHLDLSQKQTEIPVQAPSDNALKEVLL 610
Db 606 YTFKQNE-----FQNNVK-----SAKISSLESLEDIDSCKIASCNLQNVNI 652
Qy 611 PASL---KTIREAFKQNLH---KQLEVASALSHIAFNALDNDGDQFDNKNVVKTHN 664
Db 653 IASLLYPNNIQNPNKPNLLPFKQFQELKNLENVSINCILD-----694
Qy 665 SYALADGGEFIVDPKLSSTIVDLSEKILKLEGLDYSLRQTTQ-----TQPRD 713
Db 695 CHILNSISEFLEKKNKIKAFI--LKRYLLQYLDYTKLFTLQQLPELNQVYINQOLEE 752
Qy 714 MTACKALLSKNLQGEKQK-----FLQEAQFFLGRVLDLKAIA---KAERKALYTK 762
Db 753 LTVSE-----VHKQWENHKQKAFYPLCEFIKESQTLQIDFDQNTVSDSIRKILSEI 808
Qy 763 KATNGQLLERSINKAVLAYNNSAIKANVRLEKELEDLLTGLVBGKPLAQATWQGVY 822
Db 809 SESSKHYHLRLNPSQ-----SSSLIKSEN-----BEIQELLKACDEKGVLYKA-----851
Qy 823 LKTPLEPIPEYIIGLVNVPD 842
Db 852 YKFPCLP-----TGTYD 866

RESULT 54
US-08-854-050-54
; Sequence 54, Application US/08854050
; Patent No. 6261836
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6261836el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,050
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
```

```
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; US-08-854-050-54

Query Match 3.3%; Score 167; DB 3; Length 872;
Best Local Similarity 20.5%; Pred. No. 0.0036;
Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;

Qy 13 TTVSVVTHNQVFP-----SLVKEPILK---QTOASSISIGADYAESGSKLK 57
Db 17 TNLDFVLQNLVEYKSIQIEHYKTQQQIKEDLKLKFNQDQNGNSGDDDEENSH-K 75
Qy 58 INETSGPVDDTVTDLFSDKRTTPEKID-----NLAKGPREQELKAVTENTSEKQIT 110
Db 76 QQELLRRVNOIKQVQLIKKVGSKVEDLNLENKNGKGLSEQQVKEEQRLTITEQVK 135
Qy 111 SGQLEQSKESLSLNKTVPTSTNWEICDPTFKGNTLVGLSKSGVKEKLSQTHLVLPQAA 170
Db 136 YQNLVFNNDYQDLNLSGSHRRHRTDYDT-----EKWFEISH-----174
Qy 171 DGTOLIOVASPAFTPKKTAIAEYTSRAGENGEISQLDVGCKEINEGE--VFNSYLLKK 228
Db 175 --DQKNYYSIYA---NOKTSYCWMLKDYFNQNNYDHLNVSINRLETEAFYAFDFQSI 229
Qy 229 VTITGYKHIGQDAFVDNK-----NIAEV-----NLPESET 260
Db 230 KLTNNSYQTVNIDVNFNNLCILALLRFLSLERFNILNIRSSYTRNQYFEKIGELLE 289
Qy 261 ISDYAFALHAKQIDLPDLNKAIGELAPFDNQITGKLS-----298
Db 290 IFAVVFSHRHLQGIHLQVPCFAQYLVNSSQISVKDSQLQVYSFSTDCLKLVDTNKKVDY 349
Qy 299 -----LPRQLMRLAERAFKSNHIKTIERGNSLKVIGEAS-----PQDNDL 339
Db 350 PKFQEBPR-LTHVSQQAI PVSATNAVENLVLLKVKHNLNLSVPTQNFNPFYFVNL 408
Qy 340 SQLMLPDGLE-----KIESEAFPGDDHNNRVVLMTKSGKNPSGLATE-NTYVNP 391
Db 409 QHLKLEFGLEPNILTQKLEN-----LLLSIKQSKNLKFLRLNFYTYVAQ 453
Qy 392 DKSLWQSPEDYTKWLEEDFTYQKNVSGTFSNKGLOKVKKNKLEPKQKHNGVTITEIG 451
Db 454 ETSRKQ-----ILQATT-----IKLNKNNKQETPE---TKDTP 487
Qy 452 DNAFNVDQNKTLRYDLE--VKLPST---IRKIGAFQSNLKSFEAS--DDLBEI 504
Db 488 SESTGKMFDFHLSLTELDFSVNLQATQEIYDSLHLKLLIRSTNLKPKLSYKYEKMS 547
Qy 505 KEGAF--MNNRIETL-ELKDKLVTTGDAAFHINHIAIVLPESV-----QBI--GR 550
Db 548 KMDTFIDLKNIYETLNNLKRCVSNISNP--HGNISYELTNKDSYFKFKLTINQBLQAH 605
```



QY 763 KATNGQLLERSINKAVLAYNNSAIKANVKELEKELDLTGLVEGKPLAQATWVGQVY 822  
 Db 809 SESKHHYLRNPSQ-----SSSLIKSEN-----BEIQELLKACDEKGVLVKA----- 851  
 QY 823 LLKTPPLPEYIYGLNVYFD 842  
 Db 852 YYKFPCLP-----TGTYYD 866

## RESULT 56

US-09-430-323-54  
 ; Sequence 54, Application US/09430323  
 ; Patent No. 6309867

## GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.  
 Lingner, Joachim  
 Nakamura, Toru  
 Chapman, Karen B.  
 Morin, Gregg B.  
 Harley, Calvin  
 Andrews, William H.  
 TITLE OF INVENTION: NO. 6309867e1 Telomerase  
 NUMBER OF SEQUENCES: 225  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, 8th Floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 94111

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/430,323  
 FILING DATE: 29-Oct-1999  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/854,050  
 FILING DATE: 09-MAY-1997  
 APPLICATION NUMBER: US 08/851,843  
 FILING DATE: 06-MAY-1997  
 APPLICATION NUMBER: US 08/846,017  
 FILING DATE: 25-APR-1997  
 APPLICATION NUMBER: US 08/844,419  
 FILING DATE: 18-APR-1997  
 APPLICATION NUMBER: US 08/724,643  
 FILING DATE: 01-OCT-1996

## ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.  
 REGISTRATION NUMBER: 36,429  
 REFERENCE/DOCKET NUMBER: 015389-00293005

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300

## INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 872 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: peptide  
 SEQUENCE DESCRIPTION: SEQ ID NO: 54:

US-09-430-323-54

Query Match 3.3%; Score 167; DB 3; Length 872;  
 Best Local Similarity 20.5%; Pred. No. 0.0036;  
 Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;  
 QY 13 TTTSVVTHNQEVF-----SLVKEPIIK-----QTQASSISGADYAESGCKSLK 57

Db 17 TNLDFVLQNLVYKSOIEHYKTQQQIKSEDLKLLKFNQDQDQNSGNDDEENSN-K 75  
 QY 58 INETSGPVDDTVTDLFSDKRTTPBKIKD-----NLAKGRPBQELKAVTENTSEKOIT 110  
 Db 76 QOELLRRVNOIKQOVQIKKVGSKVEKDLNLEDENKNGLSQEQVKEEQRLTITEQVK 135  
 QY 111 SGSOLEQSKESLSLNTKVPSTSNWEICDFITKNGTLVGLSKSGVEKLSQTDHLVLP 170  
 Db 136 YQNLVFNMDYQDLNBSGGHRRHRTDYDT-----EKMFESH----- 174  
 QY 171 DGTOLIOVASFAFTPKDKTAJAEYTSRAGENGESQLDVGKEIINEGE--VFNSYLLKK 228  
 Db 175 --DQNYYSIYA--NQKTSYCWMLKDYFNKQNYDHLNVSINRLETAEFVAFDDPSQTI 229  
 QY 229 VTIPGYKHIGQDAFVDNK-----NIAEV-----NLPESLET 260  
 Db 230 KLTNNSYQTVNIDVNFNNLCILALLRPLLSLEFNLIRSSVTRNOYNPEKIGELLE 289  
 QY 261 ISDYAFAPHALKQIDLPDNLKAIGELAFDFNOITGKLS----- 298  
 Db 290 IFAVVFSHRHLQGIHLQVPCFAQYLVNSSQISQVDSQLOVYSFSTDLKLDVTKVODY 349  
 QY 299 -----LPROLMRLAERAFKSNHIKTIEFRGNSLKVIGEAS-----FQDNDL 339  
 Db 350 PKFLOEPR-LTHVSQQAIPVSATNAVENLVLLKKVKHANLVSPQTQNFDFYFVNL 408  
 QY 340 SOLMLPDGLE-----KIESEAFGNPDGDHNNRVVMTKSGKNPGLATE-NTYVNP 391  
 Db 409 QHLKLEFGLEPNILTQKLEN-----LLLSIKQSKNLKFLALNFYTYAQ 453  
 QY 392 DKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNGLQKVKGNKLEIPKQNGVITITIG 451  
 Db 454 ETSRKQ-----ILQOATT-----IKNLKNKQEEETPE---TKDETP 487  
 QY 452 DNAFRNVDFONKTLRKVDLBE--VKLPST---IRKIGAFQSNLKSFEAS--DDLBEI 504  
 Db 488 SESTSGMKFPDHLSELTELEDFSVNLQATQBIYOSLHLLIRSTNLKFKLSYKVEKMS 547  
 QY 505 KEGAP--MNNRIETL-ELKOKLVITGDAAPHINHIYAIVLPESV-----QEI--GR 550  
 Db 548 KMDTFIDLKNIYETLNNLKRCVSNISNP--HGNISYELTNKOSTFYKPKLTNLQELQAK 605  
 QY 551 SAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDSEQKQLTPIPVQAFSDNALKEVLL 610  
 Db 606 YTFKQNE-----POFNNVK-----SAKIESSESLESLEDIDSLCKSIASCKNLQVNI 652  
 QY 611 PASL--KTIREAFKKNHL---KOLEVASALSIAFNALDDNDGDEQFDKNVVKVTHN 664  
 Db 653 IASLLYPNNIOKNFKNPKNLLFPKQFQLKNLENVSNICILD----- 694  
 QY 665 SYALADGEHFIVDPDKLSSTIVDLEKILKLEGLDYSTLRQTTQ-----TQPRD 713  
 Db 695 QHILNSISEPLEKKNKIKAFI--LKRYVLLQYLDYTKLPKTLQQLPELNQVYINQOLEE 752  
 QY 714 MTTAGKALLSKSNLRQGEKQK-----FLOEQAQFPLGRVDLDKAI--KAELVTK 762  
 Db 753 LTVSE---VHKQVWENHKQAFYEPCEFIKESSTQLQLIDPDQNTVSDDSIKKILES 808  
 QY 763 KATNGQLLERSINKAVLAYNNSAIKANVKELEKELDLTGLVEGKPLAQATWVGQVY 822  
 Db 809 SESKHHYLRNPSQ-----SSSLIKSEN-----BEIQELLKACDEKGVLVKA----- 851  
 QY 823 LLKTPPLPEYIYGLNVYFD 842  
 Db 852 YYKFPCLP-----TGTYYD 866

## RESULT 57

US-09-402-181B-221  
 ; Sequence 221, Application US/09402181B  
 ; Patent No. 6610839  
 ; GENERAL INFORMATION:





NUMBER OF SEQUENCES: 727  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/721,456  
FILING DATE: 22-NOV-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,549A  
FILING DATE: 19-NOV-1997  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
APPLICATION NUMBER: US 08/854,050  
FILING DATE: 09-MAY-1997  
APPLICATION NUMBER: US 08/911,312  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: US 08/912,951  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: US 08/915,503  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: WO PCT/US97/17618  
FILING DATE: 01-OCT-1997  
APPLICATION NUMBER: WO PCT/US97/17885  
FILING DATE: 01-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph Ted  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002610US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 221:  
SEQUENCE DESCRIPTION: SEQ ID NO: 221:  
LENGTH: 872 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 221:  
US-09-721-456-221

Query Match 3.3%; Score 167; DB 4; Length 872;  
Best Local Similarity 20.5%; Pred. No. 0.0036;  
Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;  
QY 13 TTVSVVTHNQYVF-----SLVKEPIK-----QTQASSSISGADYAESGCKSLK 57  
DB 17 TNLDFVLQNLVYKSGIEHYKTOQQQIKEDLKLKPKNQDQNGSNDDEENNSN-K 75  
QY 58 INETSGPVDDTDLFSKRTTPEKID-----NLAKGPREQBLKAVTENSEKQIT 110  
DB 76 QQELRRVNIQKQQLIKKVGSKVKEKLNEDENKKNGLSEQQVKEEQLRTITEQVK 135  
QY 111 SSGSQLEQSKESLSLNTKVPSTNWEICDFTKNTLVGLSKSGVKLSQTHLVLPQAA 170  
DB 136 YQNLVFNMDYQDLNAGESGHRHRETDYD-----EKWFEISH----- 174  
QY 171 DGTQLIQVASFAPTPKKTAIAEYTSRAGENGEISQLDVGDKIINEGE--VFNSYLLKK 228

DB 175 --DQKQVSYIA---NQKTSYCWMLKDYFNKNYDHLNVSINRLETAEFYAFDDPSOTI 229  
QY 229 VTIPGYKHQODAFVNDK-----NIAEV-----NLPESLET 260  
DB 230 KLTNNSYQTWNIDVFNDDNLCILALLRFLLSLERFNLINRSSYTRNQYNPEKIGELLET 289  
QY 261 ISDYAFALHALKQIDLPDNLKAIGELAFPDNQITGKLS----- 298  
DB 290 IPAVVFSHRHLOGIHLQVPCFAQYLVNSSQISVKDSQSQVYSFSTDLKLVDNKNQVDY 349  
QY 299 -----LPROLMRLAERAPKSNHIKTIEFRGNSLKVIGEAS-----FQNDNL 339  
DB 350 FKPLQEFPR-LTHVSQQAIPVSATNAVENLVLLKVKHANLVLSIPTQNFDFYFVNL 408  
QY 340 SQLMLPDGLE-----KIESEAFNGPDGDHNNRVVLMTKSGKNPSGLATE-NTYVNP 391  
DB 409 QHLKLEFGLNEPILTKQLEN-----LLLSIKQSKNLKFLRLNFYTYVAQ 453  
QY 392 DKSLWQSPEDITYKNLEEDFTYQKNSVTGSKGLQVKKNKLEIPKQHNQVITIEIG 451  
DB 454 ETSRKQ-----ILKQATT-----IKNLKNNKQOETPE---TKDETP 487  
QY 452 DNAFRNVDFONKTLRKVDLEB--VKLPST---IRKIGAFAPQSNLKSFEAS--DQLEEI 504  
DB 488 SESTSGMKFPDLSELTELEDFSVNLQATQEIYDSLHLKLLIRSTNLKFKLSYKEMEKS 547  
QY 505 KEGAF--MNNRIETL-ELKOKLVITIGDAAPHINHIYAVLPESV-----QEI--GR 550  
DB 548 KMDTFIDLKNIYETLNNLKRCVSNISNP--HGNISYELTNKDSIFYKPKLTLNQLQHAQ 605  
QY 551 SAFRQNGANNLIFMGSKVKTGEMAFLSNRLHLDLSEQKQLTPIPVQAFSDNALKEVLL 610  
DB 606 YTFKQNE-----QFNNVK-----SAKTESSESLESDIDSLCKSIASCKNLQNVNI 652  
QY 611 PASL---KTIREAPKKNHL---KOLEVASALSIAFNALDDNDGDEQFDNKNVVKTHN 664  
DB 653 IASLLYPNNIOKPNFKNPILLFPKQPOLKNLENVSNICILD----- 694  
QY 665 SYALADGEHFIVDPDKLSSTIVDLKLEKILKIEGLDYSTLRQTTQ-----TQPRD 713  
DB 695 QHILNSISEFLEKKNKIKAFI--LKRYVLLQYLDYTKLFTKLQQLPELNQVYINQOLEE 752  
QY 714 MTTAGKALLSKSNLRQGEKOK-----FLOEQAQFPLGRVLDLKAIA--KAERKALVTK 762  
DB 753 LTVSE---VHKQWENHKOKAFYEPLCEFIKESSTQLQLIDFDQNTVSDSDSIKKILES 808  
QY 763 KATKNGQLLERSINKAVLAYNNSAIKKNVRLKELDLTLGLVEGKGLAQATWQGVY 822  
DB 809 SESKYHYLRNLPQO-----SSSLIKSEN-----EIQELLKACDEKGLVYKA----- 851  
QY 823 LLKTPPLPEYYIGLVNVPD 842  
DB 852 YKFPPLCLP-----TGTYYD 866

RESULT 59  
US-09-766-253-8  
; Sequence 8, Application US/09766253  
; Patent No. 6808880  
; GENERAL INFORMATION:  
; APPLICANT: Cech, Thomas R.  
; Lingner, Joachim  
; Nakamura, Toru  
; Chapman, Karen B.  
; Morin, Gregg B.  
; Harley, Calvin  
; Andrews, William H.  
; TITLE OF INVENTION: No. 6808880el Telomerase  
; NUMBER OF SEQUENCES: 171  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/766,253  
FILING DATE: 19-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/846,017  
FILING DATE: 1997-04-25  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002920US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 872 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-766-253-8

Query Match 3.3%; Score 167; DB 4; Length 872;  
Best Local Similarity 20.5%; Pred. No. 0.0036;  
Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;

QY 13 TTVSVVTHQEVF-----SLVKEPIK-----QTQASSISGADYAESSGSKLK 57  
DB 17 TNDVFLQNLVYKQIEHYKTKQQQIKBEDLKLKFKKQDQDGNDDDEENNNS-K 75  
QY 58 INETSGPVDVDTVDFSDKRTTPEKID-----NLAKGPRQELKAVTENTSEKQIT 110  
DB 76 QBELLRVNIQKQVQLIKVSKVEKDLNLEDENKNGLSQVQVKEQLRTITEQVK 135  
QY 111 SGSLQESKESLNTKVPSTSNWEICFITKNTLVGLSKGVEKLSQTDHLVLPQAA 170  
DB 136 YQLVFNMDYQLDLNESGGHRRHRETVDYD-----EKWFEISH----- 174  
QY 171 DGTQLIQVAFPTPKTKTAIAEYTSRAGENGESOLDVDGKEIINEGE--VPNSVLLKK 228  
DB 175 --DQKNVSIYA---NOKTSYCWMLKDYFNKNYDHLNVSINRLEAEFYAFDDFSQT 229  
QY 229 VTPTGYKHIGQDAFVDNK-----NIAEV-----NLPSLET 260  
DB 230 KLTNNSYQTVNIDVFNDDNLCILALRFLLSLRFNLIIRSSYTRNQYNFEKIGELLET 289  
QY 261 ISDYAFALHAKQIDLPDNLKAIGELAFPDNQITKGLS-----FQNDL 339  
DB 290 IFAVFESHRLQIHLQVPCFAQYLVNSSQISVSKDSQLQVYSFSTDLKLVDTNKVDY 349  
QY 299 -----LPRQLMELASRAFSKHITIEPRGNSLKVIGAS-----FQNDL 339  
DB 350 FKFLQBFPR-LTHYSQQAIPVSATNAVENLNLVLLKKVKHANLVSIPQFNDFDFVNL 408  
QY 340 SOLMLPDGLE-----KIESEFTGNGDDHYNNRVWLTKGKNGPSGLATE-NTYVNP 391  
DB 409 QHLKLEFEGLEPNLTQKLEN-----LLSISKQSKMLKFLRLNFTYVAQ 453  
QY 392 DKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNGKLGQKVRNKNLEIPKHNGVTITEIG 451

DB 454 ETSRKQ-----ILKQATT-----IKNLKNNKNOBETPE-----TKDET 487  
QY 452 DNAFRNVDFQNTLRKYDLEE--VKLPST---IRKIGAFAFOSNNLKSPFAS--DDELEI 504  
DB 488 SESTSGMKFFDHLSELTELEDFSVNLQATQEIYDSLHKLIRSTNLKPKLSVKYEMEKS 547  
QY 505 KEGAF--MNNRIETL-ELKDKLVTTIGDAAAFHINHIYAIYVLPESV-----QEI--GR 550  
DB 548 KMDTFIDLKNIYETLNNLKRCVSNISNP--HGNISYELTNKDSFTFYKFKLTNLQELQHA 605  
QY 551 SAFRQNGANLIFMGSKVKTGEMAFPLNRLEHLDSEQKLTETPQVAFSDNALKEVLL 610  
DB 606 YTFKQNE-----QFNNVK-----SAKIESSLESLEIDSLCKSIASCNKLNQVNI 652  
QY 611 PASL---KTIREAFKQNH---KOLEVASALSHIAFNALDNDGDGEQDPNKKVVVKTTHEN 664  
DB 653 IASLLYPNNIQKPNFKPNLLFFKQFEQKNUENVSINCILD----- 694  
QY 665 SYALADGEHFIVDPDKLSSTIVDLKILKLIBGLDYSTLRQTTQ-----TQFRD 713  
DB 695 QHILANSISEFLEKKNKIKAFI--LKRYLLQVYLDYTKLFTLQQLPELNQVYINQLEE 752  
QY 714 MTACKALLSKSNLRQGEKQK-----FLOBAQFFLGRVLDLKAIA---KAKKALVTK 762  
DB 753 LTVSE-----VHKQVMENHKQAFYEPLECFIKESQTLQIDFDQNTVSDDSIKKILSSI 808  
QY 763 KATKNGQLLERSINKAVLAYNNSAIKKANVKELEKELDLTLGLVEGKGPLAQATWVQGVY 822  
DB 809 SESKTHYLRNPSQ-----SSSLIKSEN-----BEIQELLKACDEKGVLVKA----- 851  
QY 823 LKTEPLPLPEYIYGLNVYFD 842  
DB 852 YKFPFLCLP-----TGTYYD 866

RESULT 60  
US-09-766-253-54  
; Sequence 54, Application US/09766253  
; Patent No. 6808880  
; GENERAL INFORMATION:  
; APPLICANT: Cech, Thomas R.  
; Lingner, Joachim  
; Nakamura, Toru  
; Chapman, Karen B.  
; Morin, Gregg B.  
; Harley, Calvin  
; Andrews, William H.  
; TITLE OF INVENTION: No. 6808880el Telomerase  
; NUMBER OF SEQUENCES: 171  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/766,253  
; FILING DATE: 19-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/846,017  
; FILING DATE: 1997-04-25  
; APPLICATION NUMBER: US 08/724,643  
; FILING DATE: 01-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Apple, Randolph T.  
; REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002920US

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 872 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 54:

US-09-766-253-54

```
Query Match      3.3%; Score 167; DB 4; Length 872;
Best Local Similarity 20.5%; Pred. No. 0.0036;
Matches 201; Conservative 124; Mismatches 375; Indels 280; Gaps 41;

Qy 13 TTVSVVTHNQVYP-----SLVKEPIK-----QTQASSISIGADYABSSGCKSLK 57
Db 17 TNLDFVLQNLVYKQIEHYKQQQIQKEEDLKLKFNQDQSGNDDEENNSN-K 75
Qy 58 INETSGPVDDTVDLFSKRTTPEKID-----NLAKGPREQBLKAVTENTESEKQIT 110
Db 76 QOELLRRVNIQKQVQLIKKVGSKVKKOLNLEDENKKNGLSEQQVKEQLRTITEQVK 135
Qy 111 SCSLEQSKESLSLNTKTPSTSNWEICDFTKNTLVGLSKSGVEKLSQTDHLVLPQAA 170
Db 136 YQNLVFNMDYQDLNLSGSHRRHRETDYDT-----EKWFEISH----- 174
Qy 171 DDTQLQIVASFAFTPDKTAIAEYTSRAGENCEISQLDVGDKEIINEGE--VENSYLKK 228
Db 175 --DKKNYSIYA---NOKTSYCWMLKDYFNKKNYDHLNVSINRLTEAEFAFDDFSOTI 229
Qy 229 VTIPGYKHIGQDAFVDNK-----NIAEV-----NLPLESET 260
Db 230 KLTNSYQTVNIDVFNDDLNLCTALLRFLLSERFNILNIRSSTYTNQYNFEKIGELLE 289
Qy 261 ISDTAPAHALAKQIDLPNLKKAIGELAFDNDQITKLS----- 298
Db 290 IPAVFVSHRHQGIHLQVPCFAQYLVNSSSQISVKDSQLVYSPSTDLKLVDTNKVDY 349
Qy 299 -----LPRQLMLRAERAPKSNHIKTIERGNLSKVGIEAS-----FQDNDL 339
Db 350 FKLQSEFPR-LTHVSQQAIPVSAITAVENLVLLKKVKHANLNVISIPQFNDFPFVNL 408
Qy 340 SOLMLPDLG-----KIESEAFGNPGDDHYNRRVVLWTKSGKNPSGLATE-NTYVNP 391
Db 409 QHLALEFGLPILTKQLEN-----LLLSIKQSKNLKFLRLNFTYYAQ 453
Qy 392 DKSLWQESPEIDYTKWLEBEDFTYQKNSVTGPNKGLQKVRKNKLEIPKQHNQVITIEIG 451
Db 454 ETSAAKQ-----ILQOAT-----IKNLKNKQOETPE-----TKDETP 487
Qy 452 DNAFRNVDFQNTLRKYDLEE--VKLPST---IRKIGAPAFOSNNLKSFEAS--DDLEI 504
Db 488 SESTSGMFPDHLSELTEBEDFSVNLQATQEIYDSLHKLLIRSLNKKPFKUSKYKTEMS 547
Qy 505 KEGAF--MNNRIETL-ELKDKLVITIGDAAFHINHIYAVILPESV-----QBI--GR 550
Db 548 KMDTFIDLKNIYETLNNLKRCVSNISNP--HGNISYELTNKDSFTFKPLTLNQLQELHAK 605
Qy 551 SAFRONGANNLIPMGSKVITLGEMAFPLSNRLEHLDLSEQKQTEIPVQAFPSNALKEVLL 610
Db 606 YTFKQNE-----FQFNNVK-----SAKIBSSSLESLEDIDSLCKSIASCKNLQVNI 652
Qy 611 PASL---KTIREAFKQHL-----KQLEVASALSHAFNALDNDGDGEQDNKVVVKTTHN 664
Db 653 IASLYPNNIQNPKNPKNLLFFKQFQKLNLENVSINCILQ----- 694
Qy 665 SYALADGEHFIVDPKLSSTIVDLSEKILKIEGLDYSLTRQTTQ-----TQFRD 713
Db 695 QHILNSISEFLEKKNKIKAFI--LKRYYLQYYLDYTKLPKTLQQLPELNQVYINQOLEE 752
```

```
Qy 714 MTTAGKALLSKSNLRQGEKOK-----FLOBAQFPLGRVLDLKATA---KAEKALVTK 762
Db 753 LTVSE---VHKQYWHNHKQKAFYEPCEFIKESSTQLQLIDPONTVSDDSIKKILESI 808
Qy 763 KATNGQLLERSINKAVLAVNNSAIKXANVRKLELDLLTGLVEGKGPLAAQATMVQGVY 822
Db 809 SESKVHYLRINPSQ-----SSSLIKSEN-----EETQELLKACDEKQVLVKA----- 851
Qy 823 LLKTPLPPLPEYYIGLNVYFD 842
Db 852 YKFPPLCLP-----TGTYYD 866

RESULT 61
US-09-949-016-11433
; Sequence 11433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11433
```

Query Match 3.3%; Score 167; DB 4; Length 2733;

Best Local Similarity 18.9%; Pred. No. 0.019;

Matches 212; Conservative 181; Mismatches 412; Indels 318; Gaps 48;

```
Qy 53 KSK-LKINETSGPVDDTVDLFSKRTTPEKIKONLAKGPREQBLKAVTENTESEKQITS 111
Db 280 KSKDVKIEVLQNLDD-VQLQFSEQSTLIRSLQSLQN--KSE--VLEGAERVHRHS- 332
Qy 112 GSQLEQSKESLSLNTKTPSTSNWEICDFTKNTLVGLSKSGVEKLSOT---DHLVLP 167
Db 333 -SKVEELSQAUSQKLE-----ITQDQLLLEKKRDVETLQQTIEKQQQVTEI 380
Qy 168 QAADGTQLIQV--ASFAFTPDKTAIAEYTSRAGE----- 200
Db 381 SFSMTKMWQVNLNEEKFSLGVEIKTLKEQLNLSRAEBAKQVEEDNEVSGLKQNYDEM 440
Qy 201 --NGLSOLD-----VDGKEIIN-----EGEVNSYLLKKVT 230
Db 441 SPAGQISKELQHEFDLLKCKENQKRLQALINRKEQLQVRSLREBELANLKDSEKKE 500
Qy 231 IPTGYKHIGQ-DAFVDNKNIAEVLNPESELETISDYAFALHALKQIDLPNLKKAIGELAF 289
Db 501 IPLSETERGEVEEDENKEYSEKCVTSKQIEIYLYKQITSEKEVELQHIRKDLLE-- 556
Qy 290 DNQITGKLSLPRQLMLRAERAPKSNHIKTIERGNLSKVGIEASFDQNDLSQLMLPDGLE 349
Db 557 -----KLAEEQFOALVKQM-----NQLQDKTNQIDLL-QAETSENAQI-----IQ 597
Qy 350 KIESEAFGNPGDDHYNRRVVLWTKSGKNPSGLATENTYVNP-----DKSLWQESPEI--- 402
Db 598 KLITSNTDASDGS-----VALVKETVVISPPCTGSEHMK--PELEEK 639
Qy 403 -----DYTKWLEBEDFTYQKNSVTGFSNKGILQKVKR 432
```

Db 640 ILALEKEBQLOKQALTSRKAILKKAQEKERHLREELKQOKDDYNRLQEQDFQSKS 699  
QY 433 NKLEIPKOHNGVTITTEIGNAPRNVDFQNTLRKYDLDEKVLPTSTIRKIGAPAFOSN-- 490  
Db 700 NENIGDQLQLOQVRESIDGKLPSDDQESCSSTGLBEPLFKATEQHTQPVLESNLC 759  
QY 491 -----NLKSPFASDDLEIEKEGA-----FMNRIETLELKD 522  
Db 760 POMPSSHEDASALQGGTSVAQIAQLKEIEAEKVELELVSTSTELTKKSEVFLQEQ 819  
QY 523 LVITIG---DAAPHINHIYAIVLPESVQE-IGRAFRQNGANNLI FMGSKVKTL----- 571  
Db 820 INKQGLEISLKTVSH-EAEVHAESLQQLLESSQLQAGLEHLEPLKQLDELQKLSKK 878  
QY 572 -GEMAPLSNRLE-----HLDLSEKQOL-----TEIPVQAFS-DNALKEV----- 608  
Db 879 BEDVSLSQGLSEKAALTKIOTEIIEQEDLIKALHTQLEMQAKEHDERIKQLOVELCEM 938  
QY 609 -----LLPASLKTIREAPKQNHKLQLEVASALSHI--AFNALDDN 647  
Db 939 KQKPEBIGESRAKQIQIRKLAALIS-RKEALKENKSLQELSLARGTIERLTKSLADV 997  
QY 648 DGDQEQDNK-----VVVTHNSYALADGEHFIVDPDKLSST-----IVDL 688  
Db 998 ESQVSAQNEKQTVLGRLLALQBERDKLITEMDRSLLENQSLSSCESLKLALGLETDK 1057  
QY 689 EKILKLEGLDYSLTRQTT--QTPRDMTTAGKALLSKNLRQEKQK-----LOEA 739  
Db 1058 EKLKVEISLKSIAESTEWQEKHKLQKEYBILQSVENVSNEAERTQHVVEAVRQEK 1117  
QY 740 QPFLGRVDLDKAIK-AEKALVTTKATQNGQLLERSINKAVLAYNNNSAIKQANVKELEKE 798  
Db 1118 QELYGKLRSTANKETEKQL--QEAQEQEEMKEMKVRKFAKS-----KQOKILELEE 1169  
QY 799 LDLLTLGLVGKGLAQATVWQGVYLLKTLPLPEYIYGLNVDPKSGKLIYALDMSDTTG 858  
Db 1170 NDRLEAHPAGDTAKECHET--LLSSNASMKEBELERVKMEYETLSKQFSL-MS---- 1221  
QY 859 EGQKDAYGNPILNVDBDNGYHALAVATLADYEGLDIKTLNLSKLQSLTSIRQVPTAAVH 918  
Db 1222 --BKDSLSEVQDLKHQIEG-----NVSKQANLEAT-----KHNDQNTVTEGT---- 1264  
QY 919 RADIFQAIQAAAEAPQL-LPKPGTHSEKSSSESAN---SKD 957  
Db 1265 ----QSIPEETEODSLSNSTPTCSVSPSAKSANPAVSKD 1302

RESULT 62  
US-09-583-110-4036  
; Sequence 4036, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 4036  
; LENGTH: 1078  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-4036

Query Match 3.3%; Score 166.5; DB 4; Length 1078;  
Best Local Similarity 19.0%; Pred. No. 0.0053;

Matches 226; Conservative 172; Mismatches 431; Indels 359; Gaps 59;  
QY 4 HLKTVALTUTTVTVVTHNEVPSLVKEPIKLTQOASSISGADYAE-----SSGSKSKLKN 59  
Db 35 HMTVGAEBTTTNTTQOQS-----XEVQYQORDTKNLVENDFGQTEGSSPWTGSKAQ 88  
QY 60 ETSQVPDVTVDLFSQKRTTPEKIKDNLAKGRPREQELKAVTENTSEKQITSGSQLEQSK 119  
Db 89 GWSWVDQKNS--ADASTRVTEAKDG-----AIT-----ISSPEKLR--- 124  
QY 120 ELSLNTKTVPSSTNWEI--CDFITKGNLTVLGSKSGVSKLSQTDHLVLPQAADGTOLIQV 178  
Db 125 --AAVHRMVPPIEAKKKYKLPFKITKDNKVGIARIIIESGDKRLMNSATTSGTQDWOT 182  
QY 179 --ASPAFTPKDKTATAEYTSRAGEISQOLDVQKEI---INEGVFNSYLLKKVTIPT 233  
Db 183 IEADVSPITLDVVKIKLELFEYTG-TGTVSFKDIELVEVADQSEDSQTKQLEEKIDLP 241  
QY 234 GYKHIGQDAFVDNKNIAEVLNPESETISDYAPAHALAKQIDLPDLNKAIGELAFPDNQI 293  
Db 242 GKXHV-----FSLADYTY-----KVENPD-----VASVKNGI 268  
QY 294 TGKL-----SLPQLMLAERAPKSNHIKTIIEFGNSLKVIGEASFQD 336  
Db 269 LEPLKEGTTNVIVSKDGEVKKIPKILASVKDAYTA---RLDDMNG-----IAGNQYYD 321  
QY 337 NDLSQL-----MLPDGLEKIESEAFGTGNPGDDHYNRVVLTMTKSGKNPSGLATENT 387  
Db 322 SKNEQMAKLNOBLEKQVADSLSSISQA-----DRTYLMKFSNYKMSANLTAT 370  
QY 388 Y-----VNPQKSLMQESPEI-----DYTKWLEEDFTYQKNSVTG----- 421  
Db 371 YRKLEMAKQVTPSPSRYYQDETIVRTVDSMEWHKHVYNSEKSI VGNWMDYEIGTPRA 430  
QY 422 -----PSNKGLOKVKR--NKNLEIPKQHNGVTITEIGDNAPR-----NV 458  
Db 431 INNTLSLMEKYEFSDEIKKYTDVIEKFVPDPPEFRKTT-----DNPFKALGGNLVDMGRV 485  
QY 459 DFQNTLRKYDLEEVKLPSTIRKI-----GAFAFOSNNLKSFEASDDLEIEKEGAF 509  
Db 486 KVIAGLLRKDDQE--ISSTIRSIQVFKLVQGGEGFYQDGSY-----IDHTNVAVTGAY 537  
QY 510 MNRIETLEL-----KOKLVITIGDAAFH--INHIIYAIVLPESV---QEIGRSA 552  
Db 538 GNVLDIGLSQLLPVIQTKNPKIDKMQTM-----YHWIDKSFAPLLVNGELMDMSGRSI 593  
QY 553 FRQNGANNLI FMGSKVKTIGEMAFILSNRLEHLDSEKQKLTPEIPVQAF--SDNALKEVLLP 611  
Db 594 SRANSEGHV---AAVEVLRTGI---HRIADMSEGETKQRLQSLVKTVI QSDSYVDVF-- 643  
QY 612 ASLKTIRBEAFKKNHLKQLEVASA--LSHI--APNALDDND-----GDEQPDNKV 658  
Db 644 KNLKTYKDISLMQSLSDAGVASVPTSYLSAPFNKMDKTAMYNAREKGFGLSLFSSRTL 703  
QY 659 VKTHN-----SYALADGSEHFT-----VDPDKLSSTI-----VDLEK 690  
Db 704 NYEHNNKENKRCWYTSDBGFYLYNGDLSHYSDGYNFTVNPYKMPGTETEDAKRADSDTG 763  
QY 691 ILKLIEGLDYSLTROTTQTFQFDMTMTAGKALLSKNLRQEKQKPF--LQBAQFFLGRVDLD 749  
Db 764 VLP-SAFVGTSKLDDANATATMDFTNWNQTLTA-----HKSFWMLKDKIAFLG---- 810  
QY 750 KAIKAEKALVTYKATKNGQLLERSINKAVLAYNNNSAIKKNVVRLEKELDLTLGLVECK 809  
Db 811 ---SNIQNTSTDTAATTIDQRKLESSNPYKVVND---KEASLTQEKKDY----- 854  
QY 810 GPLAQATWQGVYLLKTLPLPEYIYGLNVYFDKSGKLIYALDMSDT-----IGEQQ 861  
Db 855 -PETQSVPLESSDSKKN-----IGY-FFPKKS-----SISMSKALQKAWKQINEQ 899  
QY 862 KD-AYGNPILNVDE-----DNEGYHALAVATLADYEGLDIKTLNLSKLQSLTSIRQVPT 914  
Db 900 SKKEVNEPLTISQAHKQNGDSGYMILIPNVDRATFNQW--IKELESSLTIENNETLQSVVD 958

QY 915 AAYHRAGI-----FOAIONAA-----ABEQLLPKPGTHS- 944  
DB 959 AKQGVYGVYKVDSDYSTINQFVLKRGVYTIKREGEYKIAYNPTEQESAPDQEVFKK 1018  
QY 945 -EKSSSSANSKDRGLQSNKPTNRGRSAILPRTGSKGSFVYGIILGY 991  
DB 1019 LEQAAQPVQNSKEKEKEEKNHSDQKN--LPQTG-BEQSILASLGF 1063

RESULT 63  
US-09-107-433-4843  
Sequence 4843 Application US/09107433  
Patent No. 6800744  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
NUMBER OF SEQUENCES: 5206  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: <Unknown>  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/ 085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4843:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1080 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...1080  
SEQUENCE DESCRIPTION: SEQ ID NO: 4843:  
US-09-107-433-4843

Query Match 3.3%; Score 166.5; DB 4; Length 1080;  
Best Local Similarity 19.0%; Pred. No. 0.0053;  
Matches 226; Conservative 172; Mismatches 431; Indels 359; Gaps 59;

QY 4 HLKVTALTITVVVTHNQEVSLVKEPILKQTQASSISGADYAE-----SSGKSKLXIN 59  
DB 37 HTWTGAEEITNTTQQSQ-----KEVYQQRDTKNLVENGDFQGTEDGSSPWTGSKAQ 90  
QY 60 ETSQGVDDTVDFSDKRTTPBKIKDNLAKGPREQELKAVTENTSEKQITSGSQLEOSK 119  
DB 91 GWSTWVDQKNSS--ADASTRVIEAKDG-----AIT-----ISSPEKLR--- 126

QY 120 BSLINKTVPTSTNWEI-CDPITKNTLVGLSKGVKLSQTDHLVLPQADGTQLIQV 178  
DB 127 --AAVHRVPIEAKKKYKLRKFKIKTNDKNVKYAKRIIEESGKDKRLMNSATTSQTKDWQT 184  
QY 179 --ASFAPTDKKTAIAEYTSRAGENGEISQLDVQKEI---INEGEVFNLSYLLKKVTP 233  
DB 185 IEADYSPTLDVDKIKLELFPYETG-TGVVSPKDIELVEVADQLSEDSQTKQLEEKIDLP 243  
QY 234 GYKHIGQDAFVDNKNIAEWNLPESLETISDYAPFAHLALQIDLPDNLKAIIGELAFDNOI 293  
DB 244 GKCHV-----FSLADTY-----KVENPD-----VASVKNKI 270  
QY 294 TGKL-----SLPQLMRLAERAPKSNHIKTIIEPRGNSLKVIGEASFQD 336  
DB 271 LEPLKGBTNTVIVSKOGKEVKKIPKILASVKDAYTA---RLDDWNG---IAGNQYD 323  
QY 337 NDLSQL-----MLPDGLEKIESEAFTCNPGDDHNNRVVLWTKSGKNPSGLATENT 387  
DB 324 SKNEQMAKLNQELSGKVADSLSSISSQA-----DRTYLWEKFSNYKMSANLTAT 372  
QY 388 Y-----VNPDKSLMQSPSEI---DYTKWLBEDFTYQKNSVTG----- 421  
DB 373 YRKLLEMAKQVTNPSSRYQDETIVRTVRDSMEWMHKHVYNSEKSI VGNMWDYEIGTPRA 432  
QY 422 -----FSNKGLOKQVR--NKMLEIPKHNGVTIIEIGDNAPR-----NV 458  
DB 433 INNTLSLMKEYFSDSEIEKKYTDVIEKFPDPEHFRKTT-----DNPFKALGNLVDMG 487  
QY 459 DFQNTLRKYDLEEVKLPSTIRKI-----GAFAPQSNLKSFEASDLEEKEGAF 509  
DB 488 KVIAGLLKQDQE---ISSTIRSEIQVFKLVDQEGEFGYDQSY-----IDHTNVAYTAY 539  
QY 510 MNNRIETLEL-----XDKLVITIGDAAPH-INHIYAIYVLPESV---OEIGBSA 552  
DB 540 GNVLDGLSQLPVIQTKNPKIDKQKQTM---YHWIDKSFAPLLVNGELMDMSGRSI 595  
QY 553 FRQNCANNLIFMGSKYKTLGEMAFLSNRLHLDISEQKLTETIPQAF-SNALKEVLLP 611  
DB 596 SRANSEGHV-----AAVEVLRI-----HRIADMSGETKQRLQSLVKTIVQSDSYVD 645  
QY 612 ASLKTIREEAFKKNHLKQLEVASA--LSHI-AFNALDDND-----GDEQDNKV 658  
DB 646 KNLTKYKDISLMQSLSDAGVASVPTSYLSAFKNKMDKTAMNAEKGFGLSLFSSRTL 705  
QY 659 VKTHN-----SYALADGEHFI-----VDPDKLSSTI-----VDLEK 690  
DB 706 NVEHMKENKRGWYTSQGMFYLYNGDLSHYSDGYMPTVNPYKMPGTTTDAKRASTDGK 765  
QY 691 ILKLEGLDYTLRQTTQTFQDMTTAGKALLSKNLQGEKQKF-LOEAQFFLGRVLD 749  
DB 766 VLP-SAFVGTSKLDDANATATMDFTNMNQTITA-----HKSWFMLKDKITAFLG----- 812  
QY 750 KAIKAEKALVTKATKNGQLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLVEGK 809  
DB 813 ---SNIQNTSTDTAATTIDQKLESSNPYKVYND---KEASLTEQEKDY----- 856  
QY 810 GPLAQATMVQGVYLLKTPLEPLPEYIYGLNVYFDKSGKLIYALDMSDT-----IGBQ 861  
DB 857 -PETQSVFLESSDSKKN-----IGY-FFFKKS-----SISMSKALQKGAWKDINEGQ 901  
QY 862 KD-AYGNPILNVE-----DNEGYHALAVATLADYEGLDIKTILNSKLSQTSIRQVPT 914  
DB 902 SDKEVENEFITSAHQKQNGDSYGYMLIPNVDRATFNQM--IKELESSIENNETLQSYVD 960  
QY 915 AAYHRAGI-----FOAIONAA-----ABEQLLPKPGTHS- 944  
DB 961 AKQGVYGVYKVDSDYSTINQFVLKRGVYTIKREGEYKIAYNPTEQESAPDQEVFKK 1020  
QY 945 -EKSSSSANSKDRGLQSNKPTNRGRSAILPRTGSKGSFVYGIILGY 991  
DB 1021 LEQAAQPVQNSKEKEKEEKNHSDQKN--LPQTG-BEQSILASLGF 1065







```
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-915-136-10

Query Match 3.3%; Score 166; DB 3; Length 2366;
Best Local Similarity 20.7%; Pred. No. 0.018;
Matches 186; Conservative 129; Mismatches 304; Indels 278; Gaps 49;

QY 20 HNOEVFSLVKEPILKQTOQASSISGADYABSSGSKLKNINETSQPVDDTVTDLFS---DK 76
Db 330 YTSEHFDMLDEEV---QSSPESVL-----ASKSDKS-----EIFSSLGDM 366
QY 77 RTTPEKIKDNL-AKGPREQLKAVTENTESE---KOITSGSQLEQSKESLSLNKTVPSTS 132
Db 367 EASPLEVKIAFNKGIINOGLISVXSYCSNLIIVKQI-----ENRYKILNNSLNPAIS 419
QY 133 NWEICDFITKGNLTAVGLSKSGVEKLSQTDHLVLPQAAQDTQLIQVASF---AFTPKKT 189
Db 420 --BNDNFNTTNTFI-----DSIMAEANADNGRFWELGKYLVRVGFDPDKT 464
QY 190 AIAEYTSRAGENGESISOLDVDGKE-----IINEGEVFNFSYLLKKVTIPTGYKHG 239
Db 465 TI-----NLSGPEAAVAYQDLLMFKEGSM-NIHLI-----494
QY 240 QDAFVNKNIAEWNLPESLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSL 299
Db 495 -EADLRNFETSKNISQSTE-----QEMASLWSFDD-----524
QY 300 PRLMLAERAPKSNHIKTIFFRGNLSKLVIGEASFDON-DLSQMLPDG---LEKIESEA 355
Db 525 ARAKAQFEE---YKRN-----FEGS-----LGE---DONLDFSQNIIVVDKEYLEKISSLA 570
QY 356 FTQNPGDHYNRNVWLTKSGKNPGLATENTYVNP--DKSLMQESPEDYTKWLEEDFT 413
Db 571 RSSERGVHY-----IVLOQDKISYEACNLFAKTPYDSVLFOKNI-----DSETA 618
QY 414 YQKNSVTGFSNKGLOKVRKNKLEIPKQNGVVTITEIGNAFNRVDFQNKTLRKYLE-- 471
Db 619 YYYNPGDG---EIQIDKYKIPSIISDRPKIKLTFIGHG---KDEFNTDIFAGFDVDSL 671
QY 472 --HVKLPSITRK--IGAPAFQSNL--KSPESADDDLEEIKEGAPMNRITETLEKDK--- 522
Db 672 STEIEAIDAKEDISPKSIIEINLLGCMNFSYINVEETYPGKLL-----LKVQDKISE 725
QY 523 ---LVTIGDAAPHIN-----HIYAVLPES---VQBIGRSAPRQ-NGANN 560
Db 726 LMFPSISQDSIIVSANQYEVRIINSEGRRELLDHSGEWINKESIIKOISSKEYISFPKFN 785
QY 561 LIFMGSKVKTLEGMFLSNRLEH-----LDLSEQQLTEIPVOAFSD---NALKEVLPL 611
Db 786 KITV--KSNLPELSTLLQBIIRNNSSDIELEKVMKLTCEINVISNIDTQIVERIEE 843
QY 612 ASLKTIREAFKQNLKQLE-VASAL-----SH-IAFNALDDNDGDQFQFNKV 658
Db 844 AKNLSDSIYIKDEFKLIBESIDALCDLKQKQNELEDHSFISFE--DISETDEGFSIRFI 901
QY 659 -----VKTHNSVALADGEHFI VDPDKLSSTIVLEKILKLI EGLDYSLRQT 706
Db 902 NKETGESIFVETKTI FSEYA-----NHITIEISKIGTIFD-----938
QY 707 TOTQFRDMTTAGKALLSKSNLROGEKQKFLQEAQFFLGRVLDLDAKAKAEKALVTYKATK 766
```

```
Db 939 -----TVNGK-LVKKVNLDTTHEVNTLNAAFFIQSLIEYNSKESLSNLSVAMKVQV 989
QY 767 NGQLLERSINKAVLAYNNSAIKANVKRLEKELDLITGLVEGKGPLAQATWVGCVYL 823
Db 990 YQALPSTGLNTITDA---AKVVELSTALDETIDLLPTLSEGL-PIL-ATIIDGVSL 1041

RESULT 67
US-08-957-310-10
; Sequence 10, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-957-310-10

Query Match 3.3%; Score 166; DB 3; Length 2366;
Best Local Similarity 20.7%; Pred. No. 0.018;
Matches 186; Conservative 129; Mismatches 304; Indels 278; Gaps 49;

QY 20 HNOEVFSLVKEPILKQTOQASSISGADYABSSGSKLKNINETSQPVDDTVTDLFS---DK 76
Db 330 YTSEHFDMLDEEV---QSSPESVL-----ASKSDKS-----EIFSSLGDM 366
QY 77 RTTPEKIKDNL-AKGPREQLKAVTENTESE---KOITSGSQLEQSKESLSLNKTVPSTS 132
Db 367 EASPLEVKIAFNKGIINOGLISVXSYCSNLIIVKQI-----ENRYKILNNSLNPAIS 419
QY 133 NWEICDFITKGNLTAVGLSKSGVEKLSQTDHLVLPQAAQDTQLIQVASF---AFTPKKT 189
Db 420 --BNDNFNTTNTFI-----DSIMAEANADNGRFWELGKYLVRVGFDPDKT 464
QY 190 AIAEYTSRAGENGESISOLDVDGKE-----IINEGEVFNFSYLLKKVTIPTGYKHG 239
Db 465 TI-----NLSGPEAAVAYQDLLMFKEGSM-NIHLI-----494
QY 240 QDAFVNKNIAEWNLPESLETISDYAFALHAKQIDLPDLNKAIGELAFPDNQITGKLSL 299
Db 495 -EADLRNFETSKNISQSTE-----QEMASLWSFDD-----524
QY 300 PRLMLAERAPKSNHIKTIFFRGNLSKLVIGEASFDON-DLSQMLPDG---LEKIESEA 355
Db 525 ARAKAQFEE---YKRN-----FEGS-----LGE---DONLDFSQNIIVVDKEYLEKISSLA 570
QY 356 FTQNPGDHYNRNVWLTKSGKNPGLATENTYVNP--DKSLMQESPEDYTKWLEEDFT 413
Db 571 RSSERGVHY-----IVLOQDKISYEACNLFAKTPYDSVLFOKNI-----DSETA 618
QY 414 YQKNSVTGFSNKGLOKVRKNKLEIPKQNGVVTITEIGNAFNRVDFQNKTLRKYLE-- 471
Db 619 YYYNPGDG---EIQIDKYKIPSIISDRPKIKLTFIGHG---KDEFNTDIFAGFDVDSL 671
QY 472 --HVKLPSITRK--IGAPAFQSNL--KSPESADDDLEEIKEGAPMNRITETLEKDK--- 522
Db 672 STEIEAIDAKEDISPKSIIEINLLGCMNFSYINVEETYPGKLL-----LKVQDKISE 725
QY 523 ---LVTIGDAAPHIN-----HIYAVLPES---VQBIGRSAPRQ-NGANN 560
Db 726 LMFPSISQDSIIVSANQYEVRIINSEGRRELLDHSGEWINKESIIKOISSKEYISFPKFN 785
QY 561 LIFMGSKVKTLEGMFLSNRLEH-----LDLSEQQLTEIPVOAFSD---NALKEVLPL 611
Db 786 KITV--KSNLPELSTLLQBIIRNNSSDIELEKVMKLTCEINVISNIDTQIVERIEE 843
QY 612 ASLKTIREAFKQNLKQLE-VASAL-----SH-IAFNALDDNDGDQFQFNKV 658
Db 844 AKNLSDSIYIKDEFKLIBESIDALCDLKQKQNELEDHSFISFE--DISETDEGFSIRFI 901
QY 659 -----VKTHNSVALADGEHFI VDPDKLSSTIVLEKILKLI EGLDYSLRQT 706
Db 902 NKETGESIFVETKTI FSEYA-----NHITIEISKIGTIFD-----938
QY 707 TOTQFRDMTTAGKALLSKSNLROGEKQKFLQEAQFFLGRVLDLDAKAKAEKALVTYKATK 766
```

Db 420 --EDNDFNTTTFI-----DSIMAEANADNGRPMELGKYLVRGFFPDVKT 464  
QY 190 AIAEYTSRAGENGESQLDVGKE-----IINEGEVFNYSLLKKVTIPTGYKHG 239  
Db 465 TI-----NLSGPEAYAAAYQDMLMFKESGM-NIHLI-----494  
QY 240 QDAFVDNKNIAEVNLPESLETISDYAPAHALAKQIDLPDLNKAIGELAFFNQITGKLSL 299  
Db 495 -EADLRNFEISKTNISQSTE-----QEMASLWSFDD-----524  
QY 300 PQLMLRAERAPKSNHIKTIEFRGNLSKVIGEASFQDN-DLSQMLPGD---LEKISEA 355  
Db 525 ARAKAQPEE--YKRN-----FEGS-----LGE---DDNLDPSQNIIVDKYLLKISSLA 570  
QY 356 FTGNGPDGHHYNNRVVLTWTKSGNPSGLATENTYVNP--DKSLWQESPEIDYTKWLEEDPT 413  
Db 571 RSSERGIYHY-----IVQLQDKISYEAAACNLFAKTPYDVSVPKQNIIE-----DSBIA 618  
QY 414 YOKNSVTGFSNKGLOKVRNKNLEIPKQHGVTITEIGDNAFRNVDFQNKTLRKYDLE--471  
Db 619 YYNPGDG-----EIQEIDKYKIPSIISDRPKIKLTFIGHG---KDEFNTDIFAGDFDVS 671  
QY 472 --EVKLPSTIRK--IGAFAPQSNL--KSFASDDLEETKEGAFMNNRIETLELKD---522  
Db 672 STEIEAADLAKEDISPKSIEINLGCNMFYSINVEETYPGKLI-----LKVDKISE 725  
QY 523 -----LVTIQDAAPHIN-----HIYAVLPES--VOEIGRSAPRO-NGANN 560  
Db 726 LMPISQSDIIVSAQYEVNRINSEGRRELLDHSGEWINKEESIIKDISKEYISPNKEN 785  
QY 561 LIFMGSKVTLCGEMAFNLRLH-----LDISEQKQLTEIPVQAFSD---NALKEVLP 611  
Db 786 KITV--KSNLPSTLLQEIERNNGSSDIELEEKVMLTECINVISNIDTQIVERIEE 843  
QY 612 ASLKTIREAPKKNHLKOLE--VASAL-----SH-IAFNALDNDGDQEQDNKV 658  
Db 844 AKNLSDSIINYIKDFEKLIESIDALCOLKQNELEDHFISE--DISETDGFSIRPI 901  
QY 659 -----VKTHNSYALADGEHFIVDPDKLSSTIVDLKILIEGLDYSTLRQT 706  
Db 902 NKEAGESIPVETEKTIPEYA---NHTEBISKIGTIFD-----938  
QY 707 TOTQFRMTTAGKALLSKNSLRQGEKQFLQBAQFGLGRVLDLDAKAIAKALVTYKATK 766  
Db 939 -----TVNGK-LYKKVNLDTTHEVNTLNAAPFIQSLIETNSKESLSNLSVAMKVQV 989  
QY 767 NSQLLERSINKAVLAYNNSAIKANVKRLEKELDLTLTGLVECKGPLAQATMVQGVYL 823  
Db 990 YAOQLFSTGLNTITDA---AKVVELVSTALDETIDILPTLSEGL-FII-ATIIDGVSL 1041

## RESULT 68

US-10-011-366-10  
Sequence 10, Application US/10011366

Patent No. 6573003

## GENERAL INFORMATION:

APPLICANT: Williams, James A.

Kink, John A.

TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES

OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE

## DISEASE

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/011.366

FILING DATE: 16-NO. 6573003-2001

CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/957.310

FILING DATE: 23-OCT-1997

APPLICATION NUMBER: US 08/329.154

FILING DATE: 24-OCT-1994

APPLICATION NUMBER: US 08/161.907

FILING DATE: 02-DEC-1993

APPLICATION NUMBER: US 07/985.321

FILING DATE: 04-DEC-1992

APPLICATION NUMBER: US 07/429.791

FILING DATE: 31-OCT-1989

## ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPND-01121

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 10:

## SEQUENCE CHARACTERISTICS:

LENGTH: 2366 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-10-011-366-10

Query Match 3.3%; Score 166; DB 4; Length 2366;

Best Local Similarity 20.7%; Pred. No. 0.018;

Matches 186; Conservative 129; Mismatches 304; Indels 278; Gaps 49;

QY 20 HNOEVSLVKEPILKQTAQSSISGADYAESGSKLNETSGPVDVTDLPS---DK 76  
Db 330 YTSFDMFLDEB---QSFESVL-----ASKSDKS-----EIFSSLGDM 366  
QY 77 RTTEPKIKDNL-AKGPREQELKAVTENTSE---KQITSGSQLESKESLNTKVPST 132  
Db 367 EASPLEVKIAFNKGIINQGLISVKDSYCSNLIVKQI-----ENRKILNLSNPAIS 419  
QY 133 NWEICDFITKQNTLVGLSKSGVEKLSQTDHLVLPQAADGTQLIQVASF---AFTPDKKT 189  
Db 420 --EDNDFNTTTFI-----DSIMAEANADNGRPMELGKYLVRGFFPDVKT 464  
QY 190 AIAEYTSRAGENGESQLDVGKE-----IINEGEVFNYSLLKKVTIPTGYKHG 239  
Db 465 TI-----NLSGPEAYAAAYQDMLMFKESGM-NIHLI-----494  
QY 240 QDAFVDNKNIAEVNLPESLETISDYAPAHALAKQIDLPDLNKAIGELAFFNQITGKLSL 299  
Db 495 -EADLRNFEISKTNISQSTE-----QEMASLWSFDD-----524  
QY 300 PQLMLRAERAPKSNHIKTIEFRGNLSKVIGEASFQDN-DLSQMLPGD---LEKISEA 355  
Db 525 ARAKAQPEE--YKRN-----FEGS-----LGE---DDNLDPSQNIIVDKYLLKISSLA 570  
QY 356 FTGNGPDGHHYNNRVVLTWTKSGNPSGLATENTYVNP--DKSLWQESPEIDYTKWLEEDPT 413  
Db 571 RSSERGIYHY-----IVQLQDKISYEAAACNLFAKTPYDVSVPKQNIIE-----DSBIA 618  
QY 414 YOKNSVTGFSNKGLOKVRNKNLEIPKQHGVTITEIGDNAFRNVDFQNKTLRKYDLE--471  
Db 619 YYNPGDG-----EIQEIDKYKIPSIISDRPKIKLTFIGHG---KDEFNTDIFAGDFDVS 671  
QY 472 --EVKLPSTIRK--IGAFAPQSNL--KSFASDDLEETKEGAFMNNRIETLELKD---522  
Db 672 STEIEAADLAKEDISPKSIEINLGCNMFYSINVEETYPGKLI-----LKVDKISE 725

QY 523 -LVTIGDAAPHIN-----HIYAIVLPES--VOEIGRSAPRQ-NGANN 560  
Db 726 LMFISQDSIIIVSAQYEVNRINSEGRRELLDHSGEWINKEESIIDISSKEYISFNPKN 785  
QY 561 LIFMGSKVTLGEMAFSLNRLEH-----LDLSEKQLTEIPVOAFSD---NALKEVLIP 611  
Db 786 KITV--KSKNLPSTLLQEIIRNNSNSDIELEEKVMLTECEINVISNIDTQIVEERIE 843  
QY 612 ASLKTIREEAFKXNHLKOLE-VASAL-----SH-IAFNALDDNDGDEQFQDNKV 658  
Db 844 AKNLTSDSINYIKDEFKLTIESIDALCDLKQONELEDSHFISFE--DISETDEGFSIRI 901  
QY 659 -----VKTHNSYALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSLRQT 706  
Db 902 NKETGESIFVETEKTIIFSEYA-----NHITEISKIKGTIFD----- 938  
QY 707 TOTQFRDMTTAGKALLSKNSRQGEKQKFLQBAQFFLGRVLDLDAKAIAKALVTYKATK 766  
Db 939 -----TVNCK-LVKKVNLDTTHEVNTLNAAFFIQSLIEYNSKESLSNLVAMKVQV 989  
QY 767 NGOLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLVEGKPLAQATMVGQVYL 823  
Db 990 YAQLFSTGLNTITDA---AKVVELVSTALDETLIDLLPTLSEGL-PII-ATIIDGVSL 1041

RESULT 69

US-09-084-517-10  
; Sequence 10, Application US/09084517  
; Patent No. 6613329  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; 'COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/084,517  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CARROLL, PETER G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: OPDH-01610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2366 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-084-517-10  
  
Query Match 3.3%; Score 166; DB 4; Length 2366;  
Best Local Similarity 20.7%; Pred No. 0.018;  
Matches 186; Conservative 129; Mismatches 304; Indels 278; Gaps 49;  
  
QY 20 HNOEVFSLVKEPILKQTOQASSISGADYAESSGSKLKINETSQVDDVTVDLFS---DK 76  
Db 330 YTEHFMDLDEEV--QSSPESVL-----ASKSDKS-----EIFSSLGDM 366  
QY 77 RTTPKIKDNL-AKPREQELKAVNTENSE---KQITSGSQLEQSKESLSLNKIVPSTS 132  
Db 367 EASPLEVIAFNSKGIINQGLISVKDSCYNLIVKQI-----ENRYKILNLSNPAIS 419  
QY 133 NWEICDFITKGNLTAVLSKSGVVEKLSQTDHLVLPQAAADGTOLIOVASE---AFTPDKKT 189  
Db 420 --EDNDFNTTTFI-----DSINAEANADNGRFMELGKYLVRVGFPFDVKT 464  
QY 190 AIARYTSRAGENGISQLDVGKE-----IINEGEVFNYSYLLKKVTIPTGYKHIG 239  
Db 465 TI-----NLSGPEAVAAAYQDILLMFKEGSM-NIHLI----- 494  
QY 240 QDAFVDNKNIAEVNLPESLETISDYAFALHALKQIDLPNLKAIGELAFPDNOITKLSL 299  
Db 495 -EADLRNFEISKTNISQSTE-----QEMASLMSFDD----- 524  
QY 300 PQLMELAEAFKSNHIKTIIEFRGNLSKVIGEASQDN-DLSQLMLPDG---LEKIESEA 355  
Db 525 ARAKAQFEE--YKRN-----PEGS-----LGE---DDNLDPSQNIIVDKKEYLEKISSLA 570  
QY 356 FTGNEPGDDHNNRVVLTGSKGNPSGLATENTYVNP--DKSIWQSPESPIDYTKWLEEDFT 413  
Db 571 RSSERGIHY-----IVQLQGDKISVEAACNLFAKTPYDSVLFOKNIE-----DSEITA 618  
QY 414 YOKNSVTGFSNGKLGKVRKNRNKLEIPKQHNGVTITEIGNAFRNVDFQNKTLRYKVDLE-- 471  
Db 619 YYNPGDG---EIQEDIKYKIPSIISDRPKIKLTFIGHG---KDEFNDFIAGFDVDSL 671  
QY 472 --EVKLPSTIRK--ICAFAPQSNL--KSFASDDLEELKEGAFWNNRIETLELKK--- 522  
Db 672 STEIEAAILAKEDISPXSIEINLLGCNNMFSYINVEETYPGKLL-----LKVKDKISE 725  
QY 523 -----LVTIGDAAPHIN-----HIYAIVLPES--VOEIGRSAPRQ-NGANN 560  
Db 726 LMFISQDSIIIVSAQYEVNRINSEGRRELLDHSGEWINKEESIIDISSKEYISFNPKN 785  
QY 561 LIFMGSKVTLGEMAFSLNRLEH-----LDLSEKQLTEIPVOAFSD---NALKEVLIP 611  
Db 786 KITV--KSKNLPSTLLQEIIRNNSNSDIELEEKVMLTECEINVISNIDTQIVEERIE 843  
QY 612 ASLKTIREEAFKXNHLKOLE-VASAL-----SH-IAFNALDDNDGDEQFQDNKV 658  
Db 844 AKNLTSDSINYIKDEFKLTIESIDALCDLKQONELEDSHFISFE--DISETDEGFSIRI 901  
QY 659 -----VKTHNSYALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSLRQT 706  
Db 902 NKETGESIFVETEKTIIFSEYA-----NHITEISKIKGTIFD----- 938  
QY 707 TOTQFRDMTTAGKALLSKNSRQGEKQKFLQBAQFFLGRVLDLDAKAIAKALVTYKATK 766  
Db 939 -----TVNCK-LVKKVNLDTTHEVNTLNAAFFIQSLIEYNSKESLSNLVAMKVQV 989  
QY 767 NGOLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLVEGKPLAQATMVGQVYL 823  
Db 990 YAQLFSTGLNTITDA---AKVVELVSTALDETLIDLLPTLSEGL-PII-ATIIDGVSL 1041

RESULT 70

US-09-248-796A-16013

Sequence 16013, Application US/09248796A

Patent No. 6747137

GENERAL INFORMATION:

APPLICANT: Keith Weinstein et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.132

CURRENT APPLICATION NUMBER: US/09/248,796A

CURRENT FILING DATE: 1999-02-12

PRIOR APPLICATION NUMBER: US 60/074,725

PRIOR FILING DATE: 1998-02-13

PRIOR APPLICATION NUMBER: US 60/096,409

PRIOR FILING DATE: 1998-08-13

NUMBER OF SEQ ID NOS: 28208

SEQ ID NO 16013

LENGTH: 862

TYPE: PR1

ORGANISM: Candida albicans

US-09-248-796A-16013

Query Match 3.3%; Score 165.5; DB 4; Length 862;

Best Local Similarity 19.4%; Pred. No. 0.0045;

Matches 179; Conservative 140; Mismatches 352; Indels 251; Gaps 39;

Qy 96 LKAVTENTESEKQITSGSQLESQSKESLSLNKTPSTSNWEICDPITKGNLTVLGSLSGVE 155

Db 64 LTIHEHNNTLEALYLLSNRLQDKLSIRIQAVVALSHFQLPFSEIGDT-GEFEDELI 121

Qy 156 KLSQDHLVLPQSAADQTQIQVASFAFTPKKTAIAEYTSRAGENGELISQ----- 206

Db 122 SSNQIQNKLNLSIQNDDSPVRRALMNLVKTQDTIPILLERARDSNSINRLVYSKIAR 181

Qy 207 -----LDVQGEKII-----NEGEVFNYSLLKKVTIPTGYKHIGQD--AFVDNKVIA 250

Db 182 ELITDLDLEFEDREFFLKWLGLNDRDETVAATAATMLTI-YWYQSVNEDLLELIDQLNVK 240

Qy 251 EV-----NLPESLETI-----SDYAPAHLLKQIDLPNLKAIGLAF 288

Db 241 SATAEQAILAFFKKKPEVLSTIKIDESYKWLTKTEKAPLMRTFYQCNQLHALMDANF 300

Qy 289 ---FDNQITGKLSLPROLMRLAERAPKSNHKTIEPRGNSLKVIGEASPO-DNDLSOL-- 342

Db 301 PELLDLISITLEKVLVSRLKTINE---NENLVKTWETYNAKIDELDNQIFSLNQISRINT 357

Qy 343 ---MLPDLGKIKISEARTGNPGDDHNNRVVLWTKSGKNPFGSLATENTYVNPDKSLWOES 399

Db 358 DADNFRKLSNIEEDITEINIAKDLFKKRI---KQLKNNSG-----NLEDLITEEN 405

Qy 400 PEI\*DYTK-WLEEDFTYQKNSVTGFSNKGQLQVKRNKNL-EIPKQHNQVITITEIGDNAPR 456

Db 406 QETADQIKDFMEDLQOQLEDI-----NKNLDEIEHQPEDIT----- 442

Qy 457 NVDPQNKTRKRYDLEEVKLPSTIRKIGAFAPQSNLNKLSFEASDDLEBEIKEGAFMNNRIET 516

Db 443 -----AKLEBELQTKYDSCIR-----ALETTSEL-----KIQT 469

Qy 517 LELKDKLVTTGDAAFHNHIIYAVLPESVQIEGRSAPRQNGANNLIPMGSKVYTLGEMAF 576

Db 470 VQ-----PEQAHENDCI---PFVDALKLELEF 493

Qy 577 LSNRL-----EHLDLSEQKQLETPVQAPSDNALKEVLLPASLKTITREAFKQNLKQ 629

Db 494 IINQLLIVKDFYGDGEMARKLLHIITMTLTEDKLPDALISVALRVLRALS-----INE 548

Qy 630 LEVASALSHIAFNALDNDGDGEFDNKVVVKVTHNSYALADGEHFI VDPDKLSSTIVDLE 689

Db 549 KDFVSMAVEIITDIRDSRD-DEEFHSAATFDDDDDDILGNED----DESOQSSSLSAVT 603

Qy 690 KILKLEGL--DYSTLRQTQTQTF--RDMTTAGALLSKSNLRQO-----EKOKFL 736

Db 604 KGRIEPMDPPDDIVLRCMTQVLEVI THSLDDHLSLSIYSGIYVNYAIONESKKLY 663

Db 677 MENEKVLVACEDVRHOLEECLAGNNQSL--EKNTIV---ETLMEKEIEAEELCWAKRL 732  
QY LEEDFTYQK-----NSVTGFSNKGLOKVRKNLEIPKQHGVTITEIGDNAP 455  
Db 733 LEEANKYKTIIEBLSNARNLNTSALQLEHHLIKINQKDMELAEKKNI---EQMDTTH 789  
QY 456 RVNDFONTLRKYDLEBK-LPSTIRKIGAFAPQSNLKSFEASDDLEIRIKEGAFNNRI 514  
Db 790 K-----ETKDLVSSLEQKLTQINKKEIF-TEKLKERSKLOEELDKYSQ-ALRKNBI 843  
QY 515 --ETLEBKDLVTIGDAAHINHIYAILVPESVQIEGRSAFRQNGANNLIFMGSKVKTIG 572  
Db 844 LRTIIEKOR--SLGSKGNNH-----LQEELELRERESQRTAPVADPKTLDVTELAS 896  
QY 573 EMAPLSNRLHLD-----LSEOK-----QLTEIPVQ 598  
Db 897 EYVQLNTIKEHEEIKHKHIIEDQNSQWQLLSLOEQKEMDFRYQHEGMNATHQ 956  
QY 599 AP--SNALKEVLLPASLKTIRREAPKKNHLKQEVASALSHTAFNALDNDGDEQDNK 656  
Db 957 LFLEKDEEIKS--LQKTIEQIKTLHEERQDIQTDNSDIFQETKQVQSLNIENGSEKHD-- 1012  
QY 657 VVVKTHNSVALADGEHFIVDPKLSSTIVDLKILKLTIEGLDYSTLRQTTQTFQDMTT 716  
Db 1013 -----LSKAE-----TERLVKIGIKERELEIKLLNEKNISLTQIDQ----- 1048  
QY 717 AGKALISK--SNLROGEKQFLQEAQFFLGRV-----DLDKAIAKAEKAL 759  
Db 1049 -----LSKDEVGKLTQIQKDL-EIQALHARISSTSHQTDVVYLOQLOQVAMEREKVF 1102  
QY 760 -VTKATKNGQLLERSINKAVLAYNNSAIKKANVRKLEKELDLTLGLVEGKGFLAQATMV 818  
Db 1103 AVLNEKTRNSHLKTEYKWM---DIVAAKEAALIKLDENKKLSTRFESSGQDMFRET 1159  
QY 819 QGVYLLKTPLEPYIYGLNVPYDKSGKLIYALDMSDTIGEG-----QKDAY 865  
Db 1160 QNLSRI-----IREKDIEDALSQKQCTLLAVLQTSSTGNEAGVNSHOFELLOBRDKL 1214  
QY 866 GNPILNVD-----DNEGVHALAVATLADVEGLDITILNSKLSQLT 907  
Db 1215 KOQVKOMEWKOQVMTVQNMQHESALQBELHQLAQVILVSDN-----NSKLQ----- 1264  
QY 908 SIRQVPTAAYHRAGIFQA-----IQNAAAEALQPKPGTHSEKSSSESANSKDRGL 960  
Db 1265 -----VDYTGLIQSYEQNETKLNFGQELAQV-----QHSIQLCNTKDLIL 1306  
QY 961 QSNPKTNRGHSAILPRTGS 980  
Db 1307 -----GKLDIISPOLSS 1318

RESULT 72  
US-09-949-016-7404  
; Sequence 7404, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7404  
; LENGTH: 2047  
; TYPE: PR

; ORGANISM: Human  
US-09-949-016-7404  
Query Match 3.2%; Score 164.5; DB 4; Length 2047;  
Best Local Similarity 18.9%; Pred. No. 0.019;  
Matches 193; Conservative 173; Mismatches 389; Indels 267; Gaps 46;  
QY 29 KEPILKQTOASSISGADYAESGSKL-----KINETSGPVDVTVDLSDKRTTPEKIK 84  
Db 442 KERILAQASAEVEVPFLQOALSDAENINRLSLAQDNLK--LKMRLVLEKEK 499  
QY 85 DNLAKGPRQELKAVTENTESE---KQITSGSQSKESLSLNTKTPVTSNWEICDFIT 141  
Db 500 SLLSQEKEBLQMLKLANEYEVKSTATRDLSLSELHDLRLN---EAKQELNQSIS 556  
QY 142 KGNLVLGSKSGVEKLSQ-----TDHLV-----LPSQADGTQLQVASFATPDKKTAI 191  
Db 557 EKETLI-----ABIEBLDRQNEATHMLIKQOLSKQNEGDSIISKQLDNDEK--- 609  
QY 192 AEYTSRAGENISQLDVDGKEIINEGEVFNYSYLLKCVTIPTGYKHIGQDAFVDNKNIAE 251  
Db 610 -----RVHQLEDDKMDITKELDVQKEKLIQ-----SEVALNDLHLTK 646  
QY 252 VNLPSLETISY-----AFHALKAKQIDLPNKA-IGELAFDQNTGKLSLPRQLM 304  
Db 647 QKLEKVENLVDQLNKSQESNVSIOKENLELKEHIRQNEEELSIRNE-----LMQSLN 700  
QY 305 RLAEAFKSNHITKTIEFRGNSLKVIGEASQNDLSQL--MLPDGLEKIESEAFGTGPGDD 363  
Db 701 QDSNSNFKDTLLKEREAEVRNLK-----QNLSEQLNENLKKVADV-----K 744  
QY 364 HYNVRVWLMTK-----SGKNPSGLATENTVYNDKSLWQSPED-----YTKW 407  
Db 745 MENEKVLVACEVVRHOLEECLAGNNQSL--EKNTIV---ETLMEKEIEAEELCWAKRL 800  
QY 408 LEEDFTYQKSVTGFNSKGLQKVRKNK---LEIPKQHGVTITEIGDNAFNRVDFQNTK 464  
Db 801 LEEANKYK--TIEELSN-----ARNLNTSALQLEHEH-----LIKL 835  
QY 465 LKDYLEBEVKLPSTIRKIGAFAPQSNLKSFEASDDLEIRIKEGAFNNRIETL--ELKOK 522  
Db 836 NOKKMEIAELKKNIEQMDTDHKEKTV-----LSSSLEEQKLTQINKKEIFIEKLER 891  
QY 523 LVTIGD-----AAPHINHIYAILVPESVQIEGRSAFRQNGANNLIFMGSKVKTIGEMAP 576  
Db 892 SSKLEELDKTSQALRKNEI-----LRQTEE-----KORSLSMKE 928  
QY 577 LSNRL--EHLDSLSEKQKLTIEPQAFSDNALKEVLLPAS---LKTIRE--EAFKKNHLKO 629  
Db 929 ENNHLOEELERLREESQRTAPVA--DPKTLDSVTELASEVSQLNTIKEHELEBEIKHHQKI 986  
QY 630 LEVASALSHIAFNALDD--NDGDE--QFDNKVVVVKTHNSYALADGEHFIVDPKLSSTIV 586  
Db 987 IEDQNSQWQLLSLOEQKEMDFRYQHEGMNATHQTLFLEKDEE-----IK 1034  
QY 687 DLEKILKLTIEGLDYSTLRQTTQTFQDMTACKALLSKNSLRSQGEKOKFLQEAQFLGRV 746  
Db 1035 SLQKTIEQIK--TQJHEERQDIQTDNSDIFQETK--VQSLNIENGSEKHDLSKAE----- 1085  
QY 747 DLDKAIAKAEKALVTKATKNGQLLERSINKAVLAYNNSAIKKANVRKLEK--ELDLTLGL 805  
Db 1086 -----TERLVKG--IKERELEIKLANEKNISLTQ--QIDQLSKDEVGKLTQI 1128  
QY 806 VEGK-----GPLAQATMVQGVVLLKTPPL-----PEYV 834  
Db 1129 IQOKDLEIQALHARISSTSHQTDVVYLOQLOQVAMEREKVFVNLNEKTRNSHLKTEYH 1188  
QY 835 IGLNVVFDKSGKLIYALD-----MSDTIGEGQKDAQGNPILN-----VDEBNEY 879  
Db 1189 KMDIVAAKEAALIKLDENKKLSTRFESSGQDMFRETQNSRIIREKDIEDALSQK 1248  
QY 880 HALAVATLADVEGLDITILNSKLSQLTIR-----QVPTAAYHRAGIFQATQNAEAE 934



Db 1249 QTLAVLQTSSTGNEAGVNSNQPELLQERDKLQKVKNMEWKQVMTTVQNMQHESA 1308  
QY 935 QL 936  
Db 1309 QL 1310

## RESULT 73

US-09-710-279-3188  
; Sequence 3188, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PUS480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3188  
; LENGTH: 1279  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-3188

Query Match 3.2%; Score 164; DB 4; Length 1279;  
Best Local Similarity 18.0%; Pred. No. 0.01;  
Matches 184; Conservative 187; Mismatches 363; Indels 286; Gaps 48;  
QY 20 HNOEVFSLVKBPILKQTOASISGADYAESGSKKLKINETSGPVDVDTDLFSDKRTT 79  
Db 244 YNEQNTINTPTDATDEEKQKALDKLAKAGYNKVDQATNQVSDAKTEAID----T 299  
QY 80 PEKIDNKLAKPREQ-ELKAVTENTSESKQITSGSQLSKESLS-LNKTVPSTSNWEIC 137  
Db 300 ITNIQANVAKPSARVELDSKPEDLKRQINATPNATEBEKODAIQRLNGKRDVKN--LI 357  
QY 138 DFITKGNLVLGSKGVKLSQTDHLVPSQADGTQLIOVASFAFTPDKKTATIAEYTSR 197  
Db 358 NOORDNEVEQHNKLGLOEL-ETIIR-AMPTRKSDALQELQTKFISQT----- 402  
QY 198 AGEIGEISQLDVGDKEIINEGEVFNLSYLLKVKYTIPTGYKHIGQDAFVDNKNIAEYNLPES 257  
Db 403 -----ELINNKNKDATNEEKDEAKRLLE-----ISKNTIINQAQTN---- 440  
QY 258 LETISDYAFAHIALKQID-LPDKLKAIGELAFPDNQITGKLSLPLQMLRAER---APKS 313  
Db 441 -----NQVDNAKDN--GMNEIATIPATTIKTDAKTAIDKAEQVVTIING 484  
QY 314 NHIKTIERGNLSKLVIGEASFOQNDLSQLMLPDGLEKTESEAFPTGNPGDHHNVRVLTW 373  
Db 485 NNDATDEEKAERKLVEKA-----KIEAKSNITNSDTEREVN----- 521  
QY 374 KSGKNPSGLATEN-----TYVNPDKSLWQESPE-----IDYTKWLEEDFTYQKNSVTGF 422  
Db 522 --GAKTNGLEKINNTPSTQTKTNKQEIINDKAQELQIINNTPATEE---EKQEATNR 576  
QY 423 SNKGLOKVRNQLEIPKQHNQVTTTEIGDNAPRVDFQNTKRYDLEEVKLPEITIKI 482  
Db 577 VNAGLAQAQTN-----NAHSQEVNESKTSI-----ATIKSVQPNVINKPTAINSL 625  
QY 483 GAFATQSNLKSFEASDD--LEBIEGAPMNNRIETLEKOKLVITGDAA--PHINHIYA 538  
Db 626 ---TQEAANNQKTLIGDGNATDEKEAA---KQVLTQKLINEQIKHESQDNQVNDVKA 679  
QY 539 IVLPSVQBEIGRSAPFRQGANNLIPMGSKVKTGLGEMAFLSNRLHLDLSEKQLTEIPVQ 598  
Db 680 QAI-TAIKLINANAKRQDAINL-----TNLAESK---KSDIR 714

QY 599 AFSDNALKEVLLPASLKTIREEAFKKNHLKQLE--VASALSHI---AFNAL-DDN--DGD 650  
Db 715 ANQDATTTEE-----KNTAISIDDTLQAOARNNINGANTNALVDENLEDGK 759  
QY 651 EQFONKVVVTKTHNSYALAD-----GEHFTVDVDPKLSSTIVDLEKILKLTIEGLDYSLR 704  
Db 760 QKL-ORIVLSOTQTKQAKADIAQAIQOORSTIDQONQVATT-----EEQKQALERLNOETNG 814  
QY 705 QTTOTQPRDMTTAGKALLSKNSLRQGEKQKFLQBAQFPLGRVDLDKAIKAEKA--LVTK 762  
Db 815 VNDRIQ-----AALANQNVTD-EKNNILETIR-----NVEPIVIVKPKANEIIRK 858  
QY 763 KATNGCOLLER-----SINKAVLAYNNSAIKKNVKELEKELD 800  
Db 859 KAAEQTTLLINQONQDATTLEEKQIALGKLEVEKNEALNQVSAHSNDVK----- 906  
QY 801 LLTGLVEGKGPLAQATWVGVLKLTPLPLPEYVIGLN---VYFDKSGKLIYALDMSDT 856  
Db 907 ----IVENNG-IAKISEVH-----PETIKRNAKQEIQDQASQI-----DT 943  
QY 857 IGEQKDAYGNPILNVDEDNQGYHALAVATLADYEGLDIKTILNSKLSQL-----T 907  
Db 944 INANNKSTNEEKSAADRNV-----VAKID-----AINNITNATTQLVNDAKNSGNT 991  
QY 908 SIRQVPTAAVHRAGIFQAIQNAARAEQLLPKPGTHSEKSSSSSANSKORGLQSNPKTN 967  
Db 992 SISQILPSTAVTKTNALAAALASEAKKNNAIIDQ--TPNATAEEKEEANNKVDRLQEEADAN 1049

## RESULT 74

US-09-543-681A-5182  
; Sequence 5182, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 5182  
; LENGTH: 1534  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-5182

Query Match 3.2%; Score 163.5; DB 4; Length 1534;  
Best Local Similarity 20.6%; Pred. No. 0.015;  
Matches 237; Conservative 173; Mismatches 433; Indels 309; Gaps 61;  
QY 35 QTOASSISGADYAESGSKKLKINETSGPVDVDTDLFSDKRTTPEKID-NLA----- 88  
Db 99 QGEPSSLIENQSATIKTEKGD-LIRTKLVNESITPLFKEIKQEPPOSTSINIGTHPLH 157  
QY 89 --KGPREQ-----ELKAVTENT-----ESEKQ-----ITSGS 113  
Db 158 PYNSTNNFLIIILLAEKLSKDKKWFIDLDLKNNGVNVVERSLPERKKYVPSVIYSGK 217  
QY 114 QLE-QSKESLSLNTKVTSTSNWEICDFITKGNL-VGLSKSGVLEKLSQTD-----HLV 164  
Db 218 NLXYIQSNEFFNTQSDIFATN-----DLIATGNNAKVYVYRFG--HLAKWLDYSHDNAHIF 270  
QY 165 LPSQA---ADGTQLIQVASP-----AFTPKKTAATAEYTSRAGEIGEISQDQVD 210  
Db 271 YRDDVNKYNIDETKPKQIVHPIPTKKGSHVEFIVD---SASEYMKAGNR-----LVLD 322  
QY 211 GKEINEGEVFNLSYLLKKTIVPTGYKHIGQ-DAFVDNKNIA--EYNLPESLETISDYAPA 267  
Db 323 FKDSIALERKF-PFLEKEI---KKYRRISDPSTILAKNILLNANNLNLISLKKSSNLS 378



Db 1848 TS-VEFKDANGTSGSESTKITKOGTLITPAN-----GAGAAGANTANTISV--TKDGIS 1898  
Qy 859 EGQDAYGNPILNVDEDEGHALAVATLADYE-----GLDITILNSKLSQLT 907  
Db 1899 AGNK-AVTNVSLGKKFGDG-HTLANGTVADPEKHYNAYKDL/TNLDEKADNN-----1950  
Qy 908 SIRQVPTAAYHRAGIFQAIONA--AAEAQQLLPKPGTHSEKSSSESAN-----SKDRGLQ 961  
Db 1951 -----PTVADNTAATVGLRGLGWISADKTGTGP--NOEYNAQVRNANEVFKPSGNGIN 2003  
Qy 962 SNPKTNRG 969  
Db 2004 VSGKTLNG 2011  
RESULT 76  
US-09-248-796A-19313  
; Sequence 19313, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 19313  
; LENGTH: 1173  
; TYPE: PRP  
; ORGANISM: Candida albicans  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (210)  
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno  
US-09-248-796A-19313

Query Match 3.2%; Score 162.5; DB 4; Length 1173;  
Best Local Similarity 19.7%; Pred. No. 0.011;  
Matches 215; Conservative 176; Mismatches 390; Indels 313; Gaps 53;  
Qy 20 HNOEVFSLVKEPIKQTOQASSI-----SGADYAESGSKLK-----INET 61  
Db 272 HQEI-SRYKPELNKLOQESSLKQQLLEISSKQALYAKQSRFLKFNKXDRDSWLNT 330  
Qy 62 SGVDDTVTDLPDSKRTTPEKIKONLAKPRQELKAVTENTESKQITSGSQLES 121  
Db 331 ISKLKQITDKQEMRHISNEYK-----TRESSLE-----ELSESIKLNDLSNDEHIKT 381  
Qy 122 LSLNKTVPSTNWEICDFITKGN-----TLVGLSKSGVEKLSQTDHLVLPQAADGTQ 174  
Db 382 LANLKTINDSQQTQTLVDQKILWRDEIRLKSVDLSLNDLTNATNIV--NQTMRAQ 439  
Qy 175 LIQVASPAFTPDKTAIAEYTSRAGEN-----GEISQL-DVDDG-----EINNEGVFNYSY 224  
Db 440 AQGIA-----AVKQIAQLNLSDRVYGTVAELFNVDNKKYKTAARVIAGNSLPHIV 489  
Qy 225 LLKKVTITPGYKH-----GQDAFVDNKNI--AEVNLPESETISDYAFALHAKQIDLP 277  
Db 490 VDTVDVTAATIMEELIRNKAGRTVTFPLNRIDNIEVYDPDSHEN-----QCLPLI 538  
Qy 278 DNLKALGELAPFDNQTGLSLPROMLRLAE--RAPKSNHIKTIETFRGNSLKVIGEASFQ 335  
Db 539 KKLKNEQYKAINQIFGKTLVVSELLKGELSRVK-----575  
Qy 336 DNDLSQMLPDGLEKIESEAFNPGDDHNNRV--VLMTKSGKNPSGLATENTYVNPDKS 394  
Db 576 ---LSCITL-DG-DRVDTRGVLSGGVRYDYNKSRIDALKIOTRKQE---LEKT-----DRE 623

Qy 395 LMQSPSEIDYTKLEEDFTYQKNSVTGFSNGKLG-KVKENKNLEIPKOHNGVTITEIGDN 453  
Db 624 LVKVTBEIEST-----NSQLKNLNNELQNLNRDLDRLSVSKPEPIKIEUSQL--- 669  
Qy 454 AFRNVDFONTKRYDLEEVKLPSTIRKIGAFAPQSNLKSFEASDDLEIEKEGAFMNNR 513  
Db 670 -----TNKFNLDQ-----EISLKS--NLQNLQNTKNSIKVNLK 702  
Qy 514 IETLEKDLVTIGDAAFHINHIYVLPESVQIGRSAPFRONGANNLI FMGSKVKTLGE 573  
Db 703 QHELEL-----NSEFTQVLTDQEQ-----NELDELSK 729  
Qy 574 MAP-LSNREHLDLSEQQLTEIPVQAFSDNALKEVL--LPASLKTITREAFKKNHLKQL 630  
Db 730 LAIELESKLDHIVTRSSSELDTKI-----SGIESEVINNLQPKLNKRYQEQ-QKHOQOOL 782  
Qy 631 EV--ASALSHIAFNALDNDDEQFDNKKVVVKKTHNSYALADGHEFIVDPDKL--SSITV 686  
Db 783 QLOQESSKSDTKSNLEYEELQOELENL-----HIQLDTSQLNSQVVE 826  
Qy 687 DLEKILKLEGLDYSTLRQTQTQFRDMTTAGKALLSKSNLRQGEKQKFLQEAQFFLGRV 746  
Db 827 NLTKINBEINNC-QELAQAANKQKIITKINIEKELKQTNLL--NQKLIKSM-----RD 878  
Qy 747 DLDKAIK-----AEKALVTKKATK--NGOLLER--SINKAVLAY--NNSAIKKANVKRL 795  
Db 879 DANOKIRELGVLPPEAFQSEKIDQYSSDQLLSKLNGINQELTKYSHINKKAIQFNLNFR 938  
Qy 796 EKE-----LDLLTGLVEGKPLAQATWVGQVYLLKTPLPPEYVIGLVN 839  
Db 939 QKEDLMARRIDLNNKASIEENLITNLQOKNDKAIKSFQVAKSFQ-----985  
Qy 840 YFDK-----SGKLIYALDMSDTIGEGQDAYGNPILNVDEDEGHALAVATLADYEGLD 894  
Db 986 IFEKLVPRGTGNLIMQKKNNDNNVNATDDFNSD--NDDDDDDDD-----DIDNYSQVA 1036  
Qy 895 IKTILNSKLSQLTSTIRQVPTAAYHRAGI--PQATQNA-----AAEA 933  
Db 1037 ISVFSNKNDSQQRIEQLSGGSKSLCALIFATQNCDDPAPFLFDEIDSMLDTQYRISV 1096  
Qy 934 BQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKGSFVYIGLYTS 993  
Db 1097 ARLIHELNRNEDNNNE--GRSRGAQFICTTFR---PELLQLSGDK---FYGV---TF 1144  
Qy 994 VALLSLITAIKKK 1007  
Db 1145 SNKVSSVNEINKEE 1158

RESULT 77  
US-08-790-912-3  
; Sequence 3, Application US/08790912  
; Patent No. 5976542  
; GENERAL INFORMATION:  
; APPLICANT: Weiser, Jeffrey N.  
; APPLICANT: Plaut, Andrew G.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT  
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
; STREET: 1601 Market Street, 36th Floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103-2398  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/790,912



US-08-790-912-2

Query Match 3.2%; Score 162; DB 2; Length 2052;  
Best Local Similarity 18.2%; Pred. No. 0.029;  
Matches 208; Conservative 156; Mismatches 392; Indels 384; Gaps 51;

QY 30 EPILKOTQASSISGADVAESGSKLKLINETSQVDDTVT-----DLFSDKRTTPK 82  
DB 608 QPABESTTNSEKVPDTSNTGE-----VSNPSDSTTSVGSNKKPEHNDKSNSEK 661

QY 83 IKDNLAKPREQELKAVTENTSEKQITSGSQESKESLSLTKT-----VP 129  
DB 662 TVEEVPVNPNEGTVG-ISTQETKPVQPAETQNSGKIANENTGEVSNKPSDSKPPVE 720

QY 130 STSNWEICDFTK-----GNT-----LVGLSKSGVEKLSQTHLVLPQQAAG- 172  
DB 721 ESNQPEKNGTATKPSNGNTTSENGQTEPEKKLELRNVSDILYSQTNQYRQHVSLDG1 780

QY 173 -----TQLIOVASPAP-----TPDKK-----TATAEYTSRAGENGESIQ--- 206  
DB 781 PENTUTYFVKVSSAFKOVIPVASITEKRNQGVYKITAQKLOQELNKKYVDNFSF 840

QY 207 -LDVDGKEILNREGVNSY--LLKKVT-IPTGYKHIG-----QDAFV 244  
DB 841 YLDKAKE---ENTNFTSFLVKAINQNPSCYTHLAASLANEVELGPDERSYIKDTFT 897

QY 245 -----DNKNIAEVLNPSL-ETISDYAFALHALKQIDLPNLKAIGELAFDNIQTK 296  
DB 898 GRLIGEKDGKIYAIYNLKKPLPENLSGATVEXLSLKV-----AISGK 940

QY 297 LSLPRQLMLAERAFKSNHIKTIEFRGSLKLVIGASQDNDSQLMDPLDGKIESAF 356  
DB 941 -----NDIGSLANEATNGYKIKOVHVDG---VLAGER-----GVGGLTAK 979

QY 357 TGNPGDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLMQESPEDYTKMLEEDFTYQK 416  
DB 980 QSSIAESSFKGRIV-----NTYETTD-----1000

QY 417 NSVTGFSNGLQKVRKNKQLEIPKQHNGVTITEIGNAPRNVDFQNKTLRYDLBEVKLP 476  
DB 1001 -----AYNIGGLVGLTGNKSIASKATVTIS---SNTNRS-----D 1035

QY 477 STIRKIGAFQFQSNLKSFEASDDLEEKE-----GAFMNRIS-----TLBK 520  
DB 1036 QTVGSLAGLVDDQAHIQNSYABGDINNVAHFKQVAGVAGILMDRTSGEEKHAGELTNVLS 1095

QY 521 DKLVTIGDA--AFH-----INHIAIVLPESVQIGRSAPRQNG---ANNL 561  
DB 1096 DVNVTNGNAITGYHYTGKQVANTPSSKANRVFNVTL-EKDEVVSKESFEERGTMLDASQI 1154

QY 562 IFMGSKVKTGEMAFLSNRLEHLDLSEQKQLTEIPVQAFSDN-ALK-----EVLPLASLKT 616  
DB 1155 VSKAEINPL-----TLPTVEPLSTSGKSDSFKIAHYQANRALVYKNIEKLLPFYNS 1209

QY 617 -----IRERAFKKNHLKOLEVASALSHIAFNALDNDGDQFQDNKVYVKNTHNSYA- 667  
DB 1210 TIVKYGNLVKENS-----LYQKELLSAVNMKDDQVITDIVSNKQFANKLLL--HYNDHS 1263

QY 668 -----LADGEHFTVDPDKLSSTTVDLEKILKITEGLDYST- 702  
DB 1264 EKFDLKYQDFANLPEYNGNTGLTYPNQFLYDRDSI-----VKEVLPKLQLDYQSD 1317

QY 703 -LRQT-----TQTPRDMTTAGKALLSKNSLRQGEKQKFLQSAOFFLGRVLDKAI 752  
DB 1318 AIRKTLGISPEVKLTVELDQ-----FSKTKQNLGDSLKKLSADAGLAS---DNSV 1367

QY 753 AK--AERALKVTKATQNG-QLLERSIN-----KAVLAYNNSAIKKNVRLKEKLD 800  
DB 1368 TRGYLVDKLIKONKKEALLGLTYLERWPNYQVNVKOLVMVHPDFGKNGTSPDITLIE 1427

QY 801 LATGLVEGKPLAQATWQGVYLLKTLPLPEYIYIGLNVYFDKSGKLIYALDMSITIGE- 859  
DB 1428 L-----GKS-----GFNNLLAKNNVDYTGISLASQHGAT 1456

QY 860 ---CQKDAYGNPILNVDNEDNEGYHALAVATLADYEGL--DIKTIILNSKLSQTSIROVPT 914  
DB 1457 DLFSTLEHYRKVFLPNTSNNDFKSETKATYVEESKSTIEVKT-----KOGLA 1504

QY 915 AAYHRAGIFQAIQAAAEAEOL-----LPKPGTHSEKSSSESANSKORGLQSNPKTNR 968  
DB 1505 GTKSIGVYDRITSATWKYRNWVLPILTLPEERSVFVISTWSSLFGAYDRVRSDDHKAGK 1564

RESULT 79  
US-09-134-000C-5785  
; Sequence 5785, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 5785  
; LENGTH: 1282  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-5785

Query Match 3.2%; Score 161.5; DB 4; Length 1282;  
Best Local Similarity 19.2%; Pred. No. 0.015;  
Matches 211; Conservative 149; Mismatches 341; Indels 397; Gaps 54;

QY 147 VGLSKSGV-----EKLSTQ-----DHLVLPQQAADGTQLI 176  
DB 68 GVGAQSGTYYVDCGEGNANDGQSPASAWRDPKYNQTEFQPGDHVLLNAQSTWNNQLL 127

QY 177 Q-----VASPAFTPDKKTAAIAYTSRAGENGESISQLDVDGKEIINEGEVFNYSL 225  
DB 128 HPKNGTAAQKIVDFDVTNDKGETIPETTR-----PIINGGGTSTGT 172

QY 226 LKKVTIPGYKHIGQDAFVDNKNIAEVLNPSLETISDY-----A 265  
DB 173 FKRAI--SGAVOLVNOQYWDISNLEVTNTPB-LDNLEGYKKGPDQAQRAGILVLGYEQNRT 229

QY 266 FAHLALQK-----IDLDPN-----LKAIGEL-----AFPD---NQITGKLSLPRQ 302  
DB 230 FNSVTIRNNYVDVQTEYILNLSGNTATKRLKAVGGIIVLGSWFDENGWVTA-ANDHRT 288

QY 303 LMRILAERAFKSNHIKTIEF-----RGNSL-KVIGEASFQDNDSQLMLPDGLB 349  
DB 289 TTGFNDILLIENNVIORVGLGIRTKADSDTSRGNTFYKTFNSITIRNNYLEDI-AGDGI- 346

QY 350 KIESEAFQNGPDHYNRVVLWTKSGKNPSGLATENTYVNPDKSLMQ----- 397  
DB 347 -VLSEAKSGGVVEGNVAVRMCNADYGTQNYAGVWA-----MSVDDGLFQYNEVYGIKYGFN 401

QY 398 --ESPEID-----YTKMLEEDF---TYQKNSV-----TGFSN- 424  
DB 402 DAEDYDVMQSNVYQYNYSHNTTGGFLLMLSDQKNSVIRYNTSANDGGGNRRRTGKNP 461

QY 425 --KGLQKYK-----RNKNLEIPKQHN-----GVITITEIG-----DNAPRVNDFQ 461  
DB 462 RGAGGYNYKEQSIIFYWVKNDGAAMPITHNNTIIVVGDISTSLFEGEGSSDNGSVANFY 521

QY 462 NKTLLKVDLEBVKLPSTIRKIRKIGAFQFQSNLKSFEASDDLEEIKEGAFMNRITELK- 520  
DB 522 NNILYKSGTGQLKFLSN-----YPTNGTQPIERK--MYDNPEKYFKNNVWPKEJAT 571

QY 521 -----DKLVTIGDAAPHINHIAIVLPESVQIGRSAPR--QNGANNLIFMGSVKVTL 571

572 EKSGATVEKLVSSGN-IFEPKQPLEITDNPBKVKELAEQBEFTTLKPTKONVVEFTSKER-- 628  
572 GENAFSLNRLEHDLSEQ-----KQLTEI--PVQAFSDNALKEVLLPASLK-----TIREEA 621  
629 -----LRQAOFPRLKENSIPAIGLSEVNSPAEDFFGNSLKNKVLIDIGAQAQASTIEKSI 683  
622 FKQHLKQLEVASALS-----HIAFNALDDNDGQDFQDNKVVVTKHNSYALADGE 672  
684 RYQNV--LEISSATGVYPLPQVELTY-----BEVVNEEVVATGKBEFL----- 728  
673 HFIVDPKLSSTIVLEKILKLEGLDYSLRQTTOTQPRDMTTAGKALLSKSNLROGEK 732  
729 QWEAIPOEKINTAGTVE--VAATVIGLPIDAVKVTAKVSF-----EGELGEGKD 775  
733 QKFELOBAQ--FFLGRVLDK--AIKAKEKALVTKKATK-----NGOLLERSINKAVLA 781  
776 TVKLKTAQTAYVQKSDGNRAYSAGTAAISSGDVAYKYPYGVNYTGNVYALKKNASSAG 835  
782 YNN-----SAIKKANVRLEKELDLTLGLVEGKGPLAQATWQ--GVYLLKT 826  
836 YNRRYVEIDTOELKNYQSLKSANLELNVRYDAMNAGNTNDRKNTQFQVDVYGTDT 895  
827 -----PLPLPEYV-----IGLVVYFDK--SG----- 845  
896 NWSNNTITWNGPNLNVNEEFIAEQSFNTSSIMNQNTISIDISNLRKLQSGEKIP 955  
846 -KLIYALDMSDTIGEGQKDAYGNPILNVDEDEGYHA----- 881  
956 AKLSFLLAITDSRLPG-----YSDNAGFPAFSKEGAKAYQDFLTGKLTLP 1004  
882 --LAVATLA-----DYGLDIKTLNS--KLSQTSI-----RQV 912  
1005 QQLAEDSLAPKIVLSNVFVKESIEVTEAGQAPKLPKETTIFYSDGSGOREVTNVNMSV 1064  
913 PTAAYHRAGTFOAIONAA 930  
1065 PASSYQKEGIFTVVGRAA 1082

RESULT 80  
US-08-956-171E-5235  
Sequence 5235, Application US/08956171E  
Patent No. 6593114  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
Gil H. Choi  
Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash  
Michael R. Fannon  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5256  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,171E  
FILING DATE: 20-Oct-1997  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/009,861  
FILING DATE: January 5, 1996  
APPLICATION NUMBER: 08/781,986  
FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Mark J. Hyman  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB248P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (240) 314-1224  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 5235:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 886 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 5235:  
US-08-956-171E-5235

Query Match 3.2%; Score 161; DB 4; Length 886;  
Best Local Similarity 19.0%; Pred. No. 0.0097;  
Matches 201; Conservative 149; Mismatches 348; Indels 360; Gaps 51;

QY 2 KKHLKTVALTTLTVSVVTHNQEVSLVKEPILKQTOQASSISGADYAESGSKLKLINE- 60  
DB 30 KSTLGVASVIVSTLFLITSQH-----QQAQAEINTSD-----KISEN 67  
QY 61 -----TSGPVDDT-----VTDLFSDKRTTP---EKIKONLAK-----GPREQ-E 95  
DB 68 QNNNATTTPPKDWTQTPATQNPANTAKNYPAADESLSKDAIKDPALENKEHDIGPREQV 127  
QY 96 LKAVTENTESEK-----QITSGQLSQSKESLSLNTKVPSTSNWEICDFTKGTNTLVGLSK 151  
DB 128 FOLLKNNETQYHFFSFKDPADVYVTKKAEVELDINTASTWK-----KPEVYENNQK 181  
QY 152 SGVEKLSQT-----DHLVLPQAAQDGTOLIOVASPAFTPKKTAIAEYTSRAGENGESQL 207  
DB 182 LPVRLVSYSPVEDHAYIRFPVSDGTQLKIVSSQTQIDDEETNYDYTKL----- 231  
QY 208 DVDGKEIINEGEVFNISYLLKVTIPTGYKHIGQDAFVNKNIAEVLNLPESLETISDYAPA 267  
DB 232 -VFAPKIYND-----PSLVKSDTNDVAVTNDQSSSVASNGTNTNSN--- 272  
QY 268 HIALKQIDLPDLNKAIGELAFPDNQITGKLSLPRQ--LMRLAERAFKSNHIKTIIEPRGNS 325  
DB 273 -----QNIISTINN-ANNQPOATTNMSQPAQPKSSTNADQA--SSQPAHETNSNGMT 320  
QY 326 LKVICEASFQNDLSQMLP-----DGLEK---IESEAFTCN--PGD----- 362  
DB 321 NDKTNESNQ--SDVNQOQPPADESLQDAIKNPAIIDKSHHTADNRWPIDFQMKNDKGERQF 379  
QY 363 DHYNNRV---VLWTKSGK-NPSGLATENTY-----VNPDKSLWQSPRIDYTKWLEEDF 412  
DB 380 YHYASTVPEPATVIFTKTGPIELGLKTAETWKPFVEYEDGKLPVELVSYSDSK-----DY 435  
QY 413 TYQKNSVTGFSNKGLOKVRKNKLEIPKQHNGVTITTEIGDANFRNVDQNKTLRKLYDLBE 472  
DB 436 AYIREPV---SNGTREVKI-----VSSIEYGENIHEDYD-----YTLMV 471  
QY 473 VKLPSTIRKIGAFAPQSNLKSFEASDDELEKEGAFNNRIETLELKDCLVTIGDAAPH 532  
DB 472 FAQPIT-----NN-----PDDYVDE-----ETYNLQKLL-----APYH 499  
QY 533 INHIVAILVLPESVQIGRSAPRQNGANNLIPMGSKVKTLGEMAFSLNRLEHDLSEOKL 592  
DB 500 -----KAKTL-----ERQV 508  
QY 593 TEIPVQAFSDNALKEVLLPASLKTIREFAPKXHLKQL-----EVASALSIAFNA 643  
DB 509 YEL-----EKLQEKL-----PEKYAEYKKLQDTRVELADQVKSATFE-NV 551  
QY 644 LDDND--GDEQFDNKVVKVTHNSYALADGEHFIYVDPDKLSSTIVLEKILKLEGLDYS 701  
DB 552 TPTNDQLDQEAHFVWFSEENSESVMG--FVEHPFYTAT-----LNGQYV 598





APPLICANT: Kupke, Thomas  
APPLICANT: Gotz, Friedrich  
APPLICANT: Kempter, Christoph  
APPLICANT: Jung, Gunther  
TITLE OF INVENTION: Oxidative Decarboxylation of Peptides  
TITLE OF INVENTION: Catalyzed by Flavoprotein EpiD  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESSES:  
ADDRESSES: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/645,193B  
FILING DATE: 13-MAY-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652.1540000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 990 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
US-08-645-193B-15

Query Match 3.2%; Score 161; DB 2; Length 990;  
Best Local Similarity 19.6%; Pred. No. 0.011;  
Matches 186; Conservative 152; Mismatches 345; Indels 268; Gaps 48;  
QY 28 VKPEILKQTAQSSS-----ISGADYAESGSKSLKINETSGPVDVDTVD----- 71  
DB 77 VKESLLKYLIRNSTRTPYCMLSGVALGFESENNIKIDSSFHKKDKVLDQWLYKLVH 136  
QY 72 -----LFSDKRTT-PEKIDNLAKGPREQELKAVTENTE 104  
DB 137 YLESDDTYKDSFVIWNOQNYIYNRLYLDNNSITENKRDVLSVKYNSILVFIHEN-- 194  
QY 105 SEKQITSGS--OLEQSKESLS-----LNKTV-----PSTSNWEICDPTKG- 143  
DB 195 SKKNITIEVLQISSKYSIENKEEVKVFQVELINKKEIIFSDLRPLENKNPLDYIINSL 254  
QY 144 ---NTLAVGLSKSGVEKLSQTDHLVLPQAAQDGTQLIQVASFAPTPDKTAIAEYTSRAGE 200  
DB 255 NPKNSLVG-----TLINISN-----EITKSKMPLG 280  
QY 201 NQGISOLDVDGKEINEGEVFNYSLLKKVTIPTGYKHIGQDAFVD-NKNIAEVLNPESL- 258  
DB 281 KGEYKYLQ-----IVN-----LMSQLFVSKNVLQI--DTVIDYSRN-----ELKQSLA 321  
QY 259 EYISDYAPA-----HLALKQI-----DLPDNLKAIGELAFDQNTQKGL 297  
DB 322 DNISEAAYITLWLLSPNHFQTKTIRNYHEFFMDKYGEQVLNKLQSLDNGFNGFYPKDSY 381  
QY 298 SLPLRMLRAER---AFKGN-HIKTIEPRGNSLKVIGEASFQNDLSQMLPLDGLKES 353  
DB 382 SFSNNIAFLKEKYLALQNNSHIEITE---NDVKLE-----KNNIVSKINAPVSTE-IYS 433  
QY 354 EAFNGPGDDHNNRV---VLTWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWL-E 409

DB 434 EYFGNSIKGYBDFAVISPILGSPNAGATGFRFTGNFIKKKNLQKQKEIVH-HYNNYME 492  
QY 410 EDPYQKSVTGFSGNKGKQKVRNK-----NLETPKHQNGVTITEIG----- 451  
DB 493 NDLEISQLEAPLNRNRNVLNNRIYNTCLNLPKSIDINDIFIGATFNKLYLYSEK 552  
QY 452 -----DNAFRNVDQFQ--KTLRKVDLEEVKLPSTIRKIG-----AFAFQSNML 492  
DB 553 HDSRIVFVSNMFNYEFGSELYKFLREISFEKTKFIQIPITESIDSLPFCPRIIYKNIL 612  
QY 493 KSFASDDLEETKEGAFNNRIETLE-----LKOKLVITGDAAFHINHI---YAVLPS 544  
DB 613 KPATWKINSEMPSETENMLNRPATIREKWHIPKDVIIAFDGNRLNLLNLDKHLILKKE 672  
QY 545 VQEIGRSAPRQ---NGANN-----LJPMGSKVKTIGEMAF--SNRLEHL-DLSEQKQL- 592  
DB 673 LKHGRIRILSFINESNNRMLEIVTPLYKTSKQESFIIPKRNHFNHNLKOWFSIH 732  
QY 593 TEIPVQAFSDNALKEVLLP-----ASLKTIREBAFKKXHLKQLEVASALSH 638  
DB 733 LSP- KTYQDNFIQDYLPPFITELKVNNFINKFFYKPKEDDFIK--LRLREDEYDQ 789  
QY 639 IAFNALDND-----GDEQFDNKV-----VKTHNSYALADGEHFIVDPDKLSSTIVDLE 689  
DB 790 IYSFIKNWKDYCLNSELYDYSIVDYVPEVRYGPHVIEDIENFFMYDSLLSINIQSE 849  
QY 690 -KI-----LKLIEGLDYSLRQTTQTF-----RDMTAG-----KALLSKSNLRQGE 731  
DB 850 FKIPKEFIVAISIDFLDLYLEINKSEKEILINNAEDLYRSNDIREYKNLLAKLTNPKND 909  
QY 732 ---KQKFLQEAQFFLGRVLDLKAJAKA-EKALVTKKATQNGQLLERSINK 777  
DB 910 YEILKKEFPNLHFLFNKISILENLKTLQKSLYTSRSRIGSFIMRCNR 960

## RESULT 83

US-09-248-796A-15026  
Sequence 15026, Application US/09248796A  
Patent No. 6747137

## GENERAL INFORMATION:

APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 15026

LENGTH: 1132  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796A-15026

Query Match 3.2%; Score 161; DB 4; Length 1132;

Best Local Similarity 20.3%; Pred. No. 0.014;

Matches 201; Conservative 149; Mismatches 369; Indels 272; Gaps 47;

QY 10 LTLTVSVVTHNOEVSLVKEPILKQTAQSSISGADYAESGSKSLKINETSGPVDVDTV 69  
DB 63 MTLSSGSKPSEVSKSLKPTITTKT-----STDYLSK---AKTKAKBKVTIKESD 113  
QY 70 TDLFSDKRTT-----PEKI-----KDNLAGPREQELKAVTENTESEKQITSGSL----- 115  
DB 114 KTNSEERKTEPIQSEQLLTDKKNKVEPNSEVNLK---DNTDDMK-ATACALGPDKN 169  
QY 116 -----EOSKESLSLNKTVPSTS-----NWEICDF-----ITKGNL 146  
DB 170 TKDNDSKSETTQPKLARSESFADTSLSPVNESDTEFNELAEIPEAKDGSVVAANVL 229



QY 810 GPLAQTMTQGVYLLKTPPLPYIYIGL-----NVPYDKS-----GKL 847  
DB 821 EELLEALLKANKETQVODLQKEIKALKEEIGNVQLEKAQQLSITSKVQELQNLKKGXE 880  
QY 848 IVALDMSDTIGEGQKDA--YGNPILNVDEDNCEY--HALAVA-----TLADYEGLDI 895  
DB 881 EQWNTWKAVLEKEKEKDLANTGKWLQDBENESLKAHVQEVQAQHNKKEASSAQFEELEI 940  
QY 896 -----KTIILSKLSQITS-----IROVPTAAHYHRAGIFQA 925  
DB 941 VLKKEKENELKLEAMLEKRESLSSKTLQLQDVODENKLFKQIEQLKQNTQQQSSFF-- 998  
QY 926 IQNAAAEABQLPKPOTHEKSSS-----SESANSKORGLOSNPKNRGRHSAL----- 974  
DB 999 -----PPHEELQ-KVTSEREKEISGLWNLDSLKD-AVEHQKKNRQOQVEAVELEAK 1051  
QY 975 -----LPRTGSKGSFVYG 987  
DB 1052 EVLKKLFPKVPSPNLSYGEWLHG 1075  
  
RESULT 85  
US-09-081-149-8  
; Sequence 8, Application US/09081149A  
; Patent No. 6506889  
; GENERAL INFORMATION:  
; APPLICANT: Sieburth, Derek  
; TITLE OF INVENTION: RAS SUPPRESSOR SUR-8  
; FILE REFERENCE: UTC-02938  
; CURRENT APPLICATION NUMBER: US/09/081,149A  
; CURRENT FILING DATE: 1998-05-19  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 8  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-081-149-8  
  
Query Match 3.2%; Score 160.5; DB 4; Length 582;  
Best Local Similarity 23.6%; Pred. No. 0.0056;  
Matches 135; Conservative 69; Mismatches 211; Indels 157; Gaps 30;  
  
QY 113 SOLEQSKESLSLNTKTPVSTSNWEICDPIFKGNTLVGLSKSGVEKLSOTDHLVLPQAAADG 172  
DB 3 SLSIGKEKSKEDPKVPKSAKE-----KEAKSGGFGKESKEKEPKTK-----GKDAKDG 53  
  
QY 173 TQIIOVA--SFAFTPD---KKTAAIAYTSRAGENGESIQDVGKEIINEGEVFNLSYL--- 225  
DB 54 KXDSSAAQPGVAFSVDTIKRPNPAPGTRKSSNAEVI-----KE-LNKCREENSMRLD 106  
  
QY 226 LKKYTIPTYGKHIGQDAFVDNKNIAEAVNIPESLETISYAPFAHLAKQI-DLPDN---LK 281  
DB 107 LSKRSI-----HI-----LPSSIKELTQLTLYLSNKLQSLPAEYGCIV 146  
  
QY 282 AIGELAFPDNQITGKLSLQRLMLAERAFKSNHKTIEFRGNSLKVIGEASFOQNDLSQ 341  
DB 147 NLTWLSLNSLT---SLPDSLNLKK-----LRLDLRHNKUREIPSVVYRLDSLT 196  
  
QY 342 LMLPDGLEKIESEAFNGPDGDDHNNRVVLTGSKGNPSGLATENTYVNPDKSLWOESPE 401  
DB 197 LVU-----RFRNRTTVEKOIKNLSKLSMLSRENKIKQLPAIGE 236  
  
QY 402 IDYTKWLEEDFYQKNSVTGPNKGLQKVRKNKLEIPKQHNGVITEIGD-NARFNVDF 460  
DB 237 L-----CNLIT-----LDVAHQLEHLPR-----EIGNCTQITNLDL 268  
  
QY 461 QNKTLRKYDLEEVKLPSTTKTGAFAP---FQSNL-----KSPFASDDEEIKEGAFMNR 513  
DB 269 QHNEL-----LDGPDITIGNUSSRLGLRYNLSAIPRSIAKCSALELN---LENN 318  
  
QY 514 IETL--ELKDKLVITIGDAAFIN--HIYAIVLPESVQIEGRSAFRQNGANNLIF-MGSKV 568

DB 319 ISTLPESLSSIVKLSNLTARNCFCQLYPVGSPSOFSTIYSLNMEHNRINKIPFGIFSRA 378  
QY 569 KTLGEMAFLSNREHLNLS-----EQKOLTEIP-----VQAFSDNALK 606  
DB 379 KVLSKLNKNDQNTSLPLDFGTWTSVMELNLTATNQTLPEDVSGLVSLVLEVLILSNLNUK 438  
QY 607 EVLLP---ASLKTIREEAFKKNHLKQL--EVA 633  
DB 439 K--LPHGLNLRKRLAELEENKLESLEPNEIA 468  
  
RESULT 86  
US-09-107-532A-5007  
; Sequence 5007, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 5007:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 956 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...956  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5007:  
US-09-107-532A-5007  
  
Query Match 3.2%; Score 160.5; DB 4; Length 956;  
Best Local Similarity 19.6%; Pred. No. 0.012;  
Matches 179; Conservative 135; Mismatches 288; Indels 313; Gaps 47;  
  
QY 18 VTHNOEVSLEKPEPTLKQTQASSISGADYAESCK-----SKLKINETSQPV 65  
DB 188 ISHYQDKYIIRDTV-----NTVVG---ASRSKGETFVVPVLDNLNLSRAKI-OSNMVV 237  
QY 66 DDTVTDLFSDKRTTPEKIKDNLAKPRBQELKAVTENTESEKQITSGSQLEQSKESLSIN 125

Db 238 NDPKGELFSASKETLEK-----RGYRVLVL-----NIDDPLEMSFN 274  
Qy 126 K---TVPSTSNWEICDFTKGNWLVG--LSKSGVEKLSQTDHLVLPQAAQTQLIQVASF 181  
Db 275 PLQLVIDSWANKDVHEASKRANTLTSLMFASQMG--TDNEFFYKSAKSAVNAIILITVEH 332  
Qy 182 AFTPD--KKTAAEYTSRAGENGEISQLD--VDGKEIINEGEVFNYSYLLKKVTIPTGYK-- 236  
Db 333 CFNNDICIEKTIWYNVAQMLNELGSLFYTPDTKTEKKNALDEVFN-----TLPOGNAK 385  
Qy 237 -HIGQAFVDNKNIAEVLNPSLETSIDYAFAPHLAKQIDLPDNLKAIGELAFDNDQITG 295  
Db 386 IQYGSTSPAGKAGSI-----LSTASQ-----GIEMFTSDLF 419  
Qy 296 KLSLPRQLMRLAERAFKSNHKTIEFRONSUKVIGE-----ASF--QDNDLSQIM----- 343  
Db 420 KLT-SKMSIDLKEIGFP---KSIQKVNT-NLVGKRISISFLRKENDVIRLIKKYRVKV 473  
Qy 344 -----LPDG-----LEKISEA-----FTGNFGDDHNNRVVLTGSKGNP 379  
Db 474 KALGCVLNFNEYLOQDMDIRYEEKSRARWRIEFPKQANSHDTSVTKKFKSGKNY 533  
Qy 380 SGLATEN---TVNPKSLWQESPEIDYTKMLEEDFTYQKNSVTGFSNKGQKVRKNL 436  
Db 534 SDLEISTLRKXYTDQPTAFVWIPDYD-----FTGNFGDDHNNRVVLTGSKGNP 560  
Qy 437 EIPKOHNGVTI-----TEIGNAFRNVDFONKTKYKVDLEBVKLPSTIRKIGAPAFOSN 490  
Db 561 --PSNHVLASIFISQLYTELASNCKQTPD--KKCFRRVH-----FLLD 599  
Qy 491 NLKSFESDDEEIEKEGAFMNNRIETLEK--DKLVTIGDAAFHINHIYAIVLPESVQEI 548  
Db 600 EFGNPAIDNDMGIMTVCLGRNMLPDLVIQSYQLETRYDKAF----- 642  
Qy 549 GRSAPFRQNGANLIPMGSKVTALGEMA-----FLSNRLEHLD-----LSEOKOLTE 594  
Db 643 --KTIKENQNHILMSNDEETIEBISKCGHKHTINRSASSKHLDTSSTVSSABQERV 700  
Qy 595 IPVQAFSDNALKEVLLPASLTKTIREAPKGNHLKOLEVASALSHAFNALDDN----- 647  
Db 701 ITVERASQLIEGEQII--LRNLHQDTKRKNIRPPI-----FNTKETNMPYRQF 749  
Qy 648 ---QDGEQFD-NKVVVKTHNSYALADGE---HEIVD-----PD 679  
Db 750 LAEDFTQDNEIDIAEHNLSQDNOI PYARFIKOLKTRLEYSINNIPISEEDYQD 809  
Qy 680 KL-----SSTIVDLKILKIEGLDYSLRQTQTQFRDMMTAGKALLS---KSNLRQGEK 732  
Db 810 YLNLGSGQOTINEEBLLKLV-----NTEVKSQDM--EGQVLSQEMKINIYLKES 857  
Qy 733 QKFLQBAOFFLGRVDLDKAI-----AKASKALVTKATNGQLE-----RSINKAVLAYNN 784  
Db 858 QK---EA-----IELVKVINGRLADENSADVIK--TINSQILDTPFTQSLINDSIATHN 907  
Qy 785 SAIKKANVKRLEKEL 799  
Db 908 VOESKQDQIMILEQQL 922

## RESULT 87

US-07-853-913-2

Sequence 2, Application US/07853913

Patent No. 5338839

GENERAL INFORMATION:

APPLICANT: McKay, Ronald D.G.

APPLICANT: Lendahl, Urban

TITLE OF INVENTION: Nestin Expression As An Indicator of

TITLE OF INVENTION: Neuroepithelial Tumors

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESS: Hamilton, Brook, Smith &amp; Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/853,913  
FILING DATE: 19920319  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/660,412  
FILING DATE: 22-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/603,803  
FILING DATE: 25-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/201,762  
FILING DATE: 02-JUN-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/180,548  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-4641AAAA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1805 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-853-913-2

Query Match 3.2%; Score 160.5; DB 1; Length 1805;

Best Local Similarity 18.0%; Pred. No. 0.03;

Matches 180; Conservative 164; Mismatches 372; Indels 283; Gaps 42;

Qy 102 NTESEKQITSGSQLE-OSKESLSLNTKVPSTSNWEICDFTKGNLTVLGSKSGVEKLSQT 160  
Db 402 NAEVRAEVLPLSLQTOAPEPLMLKATVPSS----- 433  
Qy 161 DHLVLPQAAADGTQLIQVASFAPTPDKKTAIA-----EYTSRAGENGESIQDVGKEII 215  
Db 434 --AILPELEEPGK--QQGHF---PDDLTSLATNLNPHHPHPTLEAKDGSESSRSVSIQOE 486  
Qy 216 NEGEVFNYSLLK-----VTIPTGYKHIGQAFVDNKNIAEVLNPSLETISDYAFAPHLA 270  
Db 487 DEGOIWE--LVEKEADIEVKVENSQAOKTORGSLDTEETODSOGPLOKET----- 534  
Qy 271 LKQIDLPNLKAIGELAFDNOI-----TGKLSLPRQLMRLAERAFK 312  
Db 535 -----LKAUGEPLMSLKIQNYETAGKENCNSSTEGHLGTLEGPEKEKQIPLSKLE 585  
Qy 313 SNHIKTIEFRGNLSKVIGEASFQDN---DLSQLMLPDGLEKIESEAFNGNPGDDHNNRV 369  
Db 586 EKNVESEKTLGVPLSELLGKEDTRTEQDELMSKPGTLK----- 626  
Qy 370 VLWTKSGHNPGLATENTYNNPDKSLWQESPEIDYTKWLEBDFTYQKNSVTGFSNKGLOK 429  
Db 627 -RPSLSLQKE-----SQEVVRPSPKEGNLES-----WTAFKEE--SQHPLGFFGAEDQMLER 673  
Qy 430 VKRNKNLEIPKQHNQVITTEIGDNAFRNVDPQN-----KTIRKYDLBVBK 474  
Db 674 LVEKEDQSFPRSP-----BEDQACRPLQENQEPLOYEABQOILRLLEKESQESIR 728  
Qy 475 LPSTIRKIGAFAPQSNLKS---EASDDEEIEKEGAFMNNRIETLEKOLKLVITIGDAAF 531  
Db 729 SPEEDQEAQSLQENQEPLOYEABQOILE-----RLIEKESQESLKS----- 773

QY 532 HINHIYAI VLPESVQIGRSAPRONGANNILFMGSKVKTIGEMAFLSNR-LEHLDLSEOK 590  
Db 774 -----PEENQIRIGPLEREN-QKSRLYLEENQETFPVPLESRNQRPPLASLEVEREE 822  
QY 591 QLTETIPQAFPSNALKEVLLPDLAKTIR-----EEAFKQNHLLKOLEVASALSHIAFNALDD 646  
Db 823 QRIWKPLEKVSQDLSGLS-LAEENVQPLRYLEEDDCINKSLLED-----KTHKSLGSLD 875  
QY 647 NDGD-----EOPDNKVVVKTTHNSVALADGHEHFIVDPDKLSSTIVDLE 689  
Db 876 RIGDSIIIOQESQVLSRPPEDORIV-----NHLEKESQEFSSSEEE 922  
QY 690 KIL-KLIEGLDYSTLR-----QTTQTF-RDMTTAGKAL-----LKSNI----- 727  
Db 923 QVMERSLEGENHESLSSVEKEDQMVESQLESQDSQSGKSLEDBESQTFGPLEKENAESLR 982  
QY 728 -----ROGEKQKFLQEAQFFLGRVLDLK-AIAKA-EKA-LVTXKATKNGQLLERSINKAVILA 781  
Db 983 SLAQDQEEQKLEQETQOTLRVAGNEQMAVSPPEKVDPELPKPLGNDQBIARSLSGKE--- 1039  
QY 782 YNNSAIKKNVKELEKELDLTLGLVEGKGPLAQATWQVGYLLKTPPLPEYYIGLVYF 841  
Db 1040 -NOESLVSUKEGIETVKSLEIIE-----PLETAEEDLERKSI 1079  
QY 842 DKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDEDNEGYHALAVATLADYEGLDIKILNS 901  
Db 1080 DTQEPL-WSTEVARETVEPEDEPPGSLGSVDNRE-----TLTSLKESQELSGL--G 1130  
QY 902 KLSQLTSIRQVTPAAYHRAGI-FOAJONAAAEQQLPKPGTHS-----EKSSSESAN 954  
Db 1131 KWNVETRVDSQOCLQVEEGLOEQHESLREVQKQLPSSGNOQRWEDVVEGKAVQGEAP 1190  
QY 955 SKDRGLQSPKTN---RGRHSAILPRTGSGSFVYGILG 990  
Db 1191 LATTGVGTEDKAEHLRGQGGE--EBAAEAGELLQDIVG 1227

RESULT 88  
US-09-713-273A-20  
; Sequence 20, Application US/09713273A  
; Patent No. 6620987  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Steve  
; APPLICANT: Butler, Karla  
; TITLE OF INVENTION: STARCH R1 PHOSPHORYLATION PROTEINS  
; FILE REFERENCE: BBI158 US CIP  
; CURRENT APPLICATION NUMBER: US/09/713,273A  
; PRIOR FILING DATE: 2000-11-15  
; PRIOR APPLICATION NUMBER: 60/081,143  
; PRIOR FILING DATE: 1998-04-09  
; PRIOR APPLICATION NUMBER: PCT/US99/07639  
; PRIOR FILING DATE: 1999-04-08  
; PRIOR APPLICATION NUMBER: 09/679,933  
; PRIOR FILING DATE: 2000-10-05  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 20  
; LENGTH: 1493  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-09-713-273A-20

Query Match 3.1%; Score 160; DB 4; Length 1493;  
Best Local Similarity 19.5%; Pred. No. 0.025;  
Matches 218; Conservative 164; Mismatches 358; Indels 376; Gaps 60;

QY 26 SLVKEPILKQTA-----SSISGADYAESGSKKINETSQGVDDTVTDLPSDKRTTPE 81  
Db 38 SIFHQTVLCQTVABHQBSVLSVANSKGNLFLAPT-----FRGSLCVR 87  
QY 82 KIKDNLAKG-----PREQELKAVTENTESEKQITSGSQLESQSKS-LSLSNKTVPSTS 132





QY 447 ITHIGNAFRN-----VDFQNTLRKYD--LEEYKLPSTIRKIGAFAPQSNILKS--FEAS 498  
Db 1071 IDBLDKFKNGVCSNFEN-GLRSTEGSSNNVNSINSIRE-NQFWIKINDLKAELEFIN 1128  
QY 499 DLLEELKEGAFMNRRIETLEKDKLWTIGDAAPHIN 534  
Db 1129 QERDDUKS-----EVLELKRLLENNTKQVN 1156

RESULT 91  
US-09-583-110-5243  
; Sequence 5243, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: PATH00-07A  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 5243  
; LENGTH: 1963  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-5243

Query Match 3.1%; Score 159; DB 4; Length 1963;  
Best Local Similarity 19.6%; Pred. No. 0.043;  
Matches 245; Conservative 174; Mismatches 386; Indels 444; Gaps 64;

QY 29 KEPILK-----QTOASSISGADYAESGKS-----KLKINETSGPV 65  
Db 403 KEPVDKSELNNQIDKASSVPTDYSTASYNALGPVLETAKGVYASEPVQPEVNETNKL 462  
QY 66 DDTVTDLDFDKTTPPKIK-----DNLAKPRE--OELKAVNTES-----EKOI 109  
Db 463 KTAIDALNDVKSELQBLQAEQQAQADYSAKTWREFKIAELQAEINNQTTPLPKQSEI 522  
QY 110 TSGSQ-LEQSKESLSLNTKV-----PSTSN-----WEI-----136  
Db 523 DANTKALQALQALQAVDKTVLQNAINTANSKREEEYTAQTWKALBEDALTAVNPVNEDETA 582  
QY 137 ----CDFITKG-----NTLVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVASFAPDPK 187  
Db 583 TQSKVDEATRNLEEAINNVLVTEKPVLTFTFDKALEREV-----VAKYLENQ 634  
QY 189 KTAIAEYTS--RAGENGESQLDVGKEIINEG-----EVFNSYL-----225  
Db 635 KTKIKSITATLKGET--VVSITVELIGDDVTNETITSAPKNLEYKEYTSLTWTWYDRGDG 693  
QY 226 -----LKKVTIPGYKHIGQDAFVDNKNIAEVLNLPSELETI--SDYAFALH 269  
Db 694 DVTEILDNOPIQLDLKKVEL-----KNIKRTDIKYENGKETN--ESLITVTPDDKRNYYL 747  
QY 270 ALKQIDLPNLKAIGEL-----AFFDNOITGKLSLPQLMRLAERAFKSNH 315  
Db 748 KITSKNQKTTLLAVNKIEETTVNGTPVYKVTAIADNLVS-----RTADNKFEEY 797  
QY 316 IKTIE-----FRGN--SLKVIGEASFQDND-----LPGLEKIESAFTGN 359  
Db 798 VHYIEPKVHEDNVYNNFKELVEAIQNDPSKEYRLQGSMSARNVWFGKSYITKE--FTGK 856  
QY 360 -----PGDDHYN 367  
Db 857 LLSSEKQPAITEHPLFNVTNATNNVNFENBIERSQDNTIASLANTWKGSVITN 916

QY 368 RVVLWTKSGKN-----PSGLATENTYVNPDKSLWOESPEIDYITKWLBEEDFTYQKNS 418  
Db 917 VKITGTLSGRNVAGVFNMMNDGTRIEN--VAFPGKLHSTSGSGHT-----CG 963  
QY 419 VTGFSNKGLOKV-----KRNKNLEIPKQHNGVITTE--IGNAFNRVDFQNTLRK 467  
Db 964 TAGTYRGIVRKAYVVDATTITGNKTRASLLVPKVDYGLTDLHLIGTKALLT--ESVVGK 1020  
QY 468 YDLEEVKLPSTIRKIGAFAPQSNILKSFEASDDLEIEKEG--AFMNRRIETLEKDKLVT 525  
Db 1021 IDV-----SNPEVGATASKTPWPGTVSNVSVYAKIIRGESELFSGNDVDSY-----1068  
QY 526 IGDAAFHINHIIYAILVPESVQIGRSAPFQNCANLLIF---MGSKVKTIGEMAFNSRLE 582  
Db 1069 ---ASAHIKDLVAVEGYSS---GNRSFRKSKTFTKLTKEQADAKVTTFN---ITADKLE 1118  
QY 593 HLDLSEKQLTPIPVQAFS-----DNALK--EVLPLASLKIT---REAFKKNHL 627  
Db 1119 S-DLSPLAKLNE--EKAYSSIQDYNAEYNAQYKLEKLIFFYKNDYIVYQGNKLNKEHHL 1175  
QY 628 KQLEVASALSHTAFNALDDNDGEOFD--NKVVV-----KTHNSYALADG-----EHF 674  
Db 1176 NTKVELSV-----TAMNNEFITNLDANKIIVHAGTKDYFNLSSESSEGLSNVKEYT 1229  
QY 675 IVD-----PDKLSSTIV--DLEKILKLEGLDYSTLRQTTQTQFRDMTTAGKALLSK 724  
Db 1230 ITDLGIKYTPNIVQKDNITLVNDIKSILESV-----LQSQTYQHNLRLGDYRVN- 1280  
QY 725 SNLROGEKQFLOEAQFFLGRVDDLKAKAEKALVTKATKQGLLERSINKAVLAYNN 784  
Db 1281 ----AIKDLYLEES-----FTDV-----KENLTNLTITKLQVNEEHQLN-----DS 1316  
QY 785 SAIKKANVKRLK-ELDLTLTGLVECKGPLAQATMVGQVYLLK-----TLPPLPEYVI 835  
Db 1317 PAARMIRDKVEKNKAALLGL-----TYLNRYYGVKFGDVNIKELMLFKPDFY- 1365  
QY 836 GLNVYDFSGKLIYALDMSDTITGEQK-----DAYGNPILNVDEDEGYPHALAVATL 887  
Db 1366 -----GEKVSVDRLLEIGSKENNIKGSRTFDFGQ-----VJ 1398  
QY 888 ADY-EGLDIKTILSKLSQTSIROVPTAAHYHRAGIFQAIQNAABAEQLLPKPGTHSEK 946  
Db 1399 AKYTKSGNLDAFLNTNRQLFTNIDNMDFIDATEDHVVYIAERASEVEEI-----1448  
QY 947 SSSSESSAKRGLOSNPKTRGRHSAILPRTGSKGSFVYGLGYTSVA 995  
Db 1449 -----KNSKRAF-DNLKRSHLRNT-ILPLNLIDKAHLYLISNYAIA 1489

## RESULT 92

US-09-107-433-4883

; Sequence 4883, Application US/09107433

; Patent No. 6800744

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID

; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE

; THERAPEUTICS

; NUMBER OF SEQUENCES: 5206

; CORRESPONDENCE ADDRESS:

; ADDRESS: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: &lt;Unknown&gt;

; OPERATING SYSTEM: &lt;Unknown&gt;

; SOFTWARE: &lt;Unknown&gt;

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,433

FOR DIAGN



```
QY 11 TLTVSVVTHNOBVSFLVKEPILKQOASSISGADYABSSGSKLKNINETSQPVDDTVT 70
Db 21 TLTVN-----F3FWOE-----RETSQVSAHFF-----PHLQENVS-RLSDIQ 63
QY 71 DLPSDKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQSKESESL-----SLN 125
Db 64 HLKSKR-----ERWWRASRQSEKEMELHND-----DLLQEKKSLECEVEELH 111
QY 126 KTVPTSNWEICDFTIKGNTLVGLSKSGVEK-----LSQT 160
Db 112 RTVQKQOQK--DFI--DGNVESLMTLEIEKSLKHEDIVDEIECKIEKTLKRSSELRA 168
QY 161 DHLVLSQAADGTQLIQVAFPTPK-KTAAIAYTS-----RAGENG 202
Db 169 DRLLAEAE-----SELSCTKEKTKNAVEKFTDAKRSLLQTESDAEELERRAQETA 218
QY 203 -----EISQDLDVGKEI-----INEGEV-----FNSYLLKKVTPTGYK 236
Db 219 VNLVKADQQLRSQADAKDLQEHKIQEBILKEINKIVAAKSDDFQCLSKKKEKLEELQ 278
QY 237 HIGQAFVNDKNIAEVNLPESLETISDYAFALHAKQIDLPNLKA-----IGELAFPN 291
Db 279 KIQKDI-----EWAERNEHHLOVLKE-SEVILQAKRAEL-EKLSQVTSQQQKAVLDR 331
QY 292 Q-----ITGKLSPLQMLAERAP--KSNHKTIEFRGNSLKVIGEASF 334
Db 332 QLGHKKEELHLLQGSVMQAKADL-QEALRGTEVTEKCNHIREVK-----SLLELSF 384
QY 335 QDNOL-----SOLMPDGLKEIESEAFNGDDHYNRVVLTKSGKNPGLATEN 386
Db 385 QXGELNVQISERTQTLTK--QETSEKE-----ENLQVVLROMSKHTELK--- 429
QY 387 TVYNPKLSQWSPEDYTKWLEED---FTYOKNSVTGFSNK-----GLOKVRKNKLETP 439
Db 430 ---NILDMLQENHELOGLK-LQHDORVSELEKTQAVLEEKLELENLQISQQQGEIE 485
QY 440 KQHNGVTITEIGDANFRVDFQNKTLR-----KYDLEEVKLPSTIRKIGA 484
Db 486 WQKQ---LLERDKREITERMTAESRALQSCVECLSEKEKDLQEKCDIWEKKLAQTKRVLAA 542
QY 485 -----FAPQSNLKSFEAS-----DLLEEI-KEGAFMNNRIETLE-----LKDK 522
Db 543 AEENSKMEQSNLEKLELNVRLQELDLNRDKLSLHNDI-SAMQQQLQEKREAVNSLOBE 602
QY 523 LVTIGDAAPHINHIAIVLPESVOEIGRSAPFGNGANNLI-FWGSVKVKTIGEMAFLSNLE 582
Db 603 LANVQD---HLN-----LAKQDLHTTKHOD-----VLLSEQTRLOKQISEWANRFE 646
QY 583 HLDSLQKQLTEIPV--QAFSDNALKEVLLPASLKTIREEAFKQNHKLQLEVASALSHTA 640
Db 647 DCQKEBETKQOQLVQLONEIEENKGLVQOEMMFQRLQKE--RESEESKLET----- 696
QY 641 FNALDNDGDEQDNKVVVK--THNSYALAGEHFIVDPDKLSSTIVDLKILKI----- 695
Db 697 -----SKVTLKEQHQHLEKELTDQSKSL---DOVLSKVLAABERVRTLOBE 739
QY 696 -----EGLDYSYTLRQTTQTFQDMWTAGKALLSKSNLRQGEKQKFLQEA----- 739
Db 740 ERWCESLE-KTLSQTK-----QLSREEQQLVEKS-----GELLALQKEADSMRADPSLNRN 790
QY 740 QPFLGRVDLDKAIKAEKALVTKKAFKNGQLLERSINKAVLANNNSAIKK----- 789
Db 791 QFLTERKKAQKQVASKALKIORSQLEKNLLEQK-----QENSCIQEMATIIELVNAQ 843
QY 790 ---ANVRLEKELD 800
Db 844 DNHERRARLMKELN 857
```

RESULT 94

US-08-409-995-4

; Sequence 4, Application US/0840995

; Patent No. 5646259

```
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RPT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-4
```

Query Match 3.1%; Score 158.5; DB 1; Length 1912;

Best Local Similarity 18.0%; Pred. No. 0.045;

Matches 211; Conservative 158; Mismatches 388; Indels 412; Gaps 55;

```
QY 47 YAESSGSKLKNINETSQPVDDTVTD-----LFSDKRTTPE-----KIKON 86
Db 566 HVEDAYKGLLNLNKNANKQPLVTDSTAATVGDRLKLGWVSTKNGTKESNQVKQADE 625
QY 87 LAGPREBELKAVTEN-----TESEKQITSGSQSKESESLNKTVPSTWSWEICDFT 141
Db 626 LFTGAGAATVTSKSENGKHTITVSVAETKADCGLEKGDGTIKLKVDNQNTDN---VLT 680
QY 142 KNTLVGLSKSGVE--KLSQTDH-----LVLPQAAAD-----GTOLIQVASPAF 183
Db 681 VGNNGTAVTKGFEVTKGATDADRKGVTVKDATTANDADKKVATVKDVATATNSAATFVK 740
QY 184 TPDKKTAAE-----YTSRAGENGESQLDVGK-----EINIEGEVFN 222
Db 741 TENLFTSIDENPTDNGKDDALKAGDTLTFKAGKULKVKR---DGKNITFDLAKNLEVT 797
QY 223 SYLLKKVTI-----PTGYKHIGQDAFDVKNKTAENVLPESLETISDYAFALHAKQLDLP 278
Db 798 AKVSDTLTIGGNTPTG---GTTA-----TPKNITSTADGLN---FA---KE 835
QY 279 NLKATGELAFPNQITGKLSLPRQLMRLAERAFKSNHIK----- 317
Db 836 TADASGSKNVYLGKATTLTEP-----SAGAKSHVDNLVDATKSKNAASIEDVLRAG 888
QY 318 -TIEFRGNSLKVIG---EASFQDND-----LSQMLPDQ-----LEKIESA 355
Db 889 WNIQGNNGNVYVAITYVTNVFTDDSTGTTTIVTQKADGKADGVKIGAKTSVKGHNGKL 948
QY 356 FTGNGPDHNNRVVLTKSGKNP--SGLATENTYVNP--DKSLWQSPSEIDYTKWLEEDFT 413
Db 949 FTGKDLKD--ANNGATVSEDDGKDTGTGLVTAKTVIDAVNKGWVRVTGE-----GA 997
```

```
QY 414 YQKNSVTGFSNKGKQKVRKNKLEIPKQHNGVTITEIGDAPFRVNDPQNKTLRYKDYLE-- 471
Db 998 TAETGATAVNAAGNAETVTSVGNP-KNGNATTATVSKDNGNIN--KYDVNVG 1048
QY 472 -----EVLKPSITIRKIG-----AFAPQSNLKSFEASD 499
Db 1049 DGLKIGDDKKIVADTTTLTVTGKVSVPAGANSVNNKKLVNAEGLATALLNL-SWTAKA 1107
QY 500 DL-----BEIKEG-----AFMNNRIE-----TLELKDKL-----VTIGDAAF 531
Db 1108 DKYADGESEGETDQEVKAGDKVTFKAGKKNLVKQSEKDFTVSLQDTLTGLTSITLGGTAN 1167
QY 532 HINHIYAIVLPESVQ-EIGRSAPFRQNGANNLIPMGSKVKTIGEMAFLSNRLEHLDLSEQK 590
Db 1168 GRNDTGTVINKDGLTITILANGAAAGTDASN-----GNTISVTKDGISAGN-----K 1213
QY 591 QLTEIPVQAFSDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSH-----I 639
Db 1214 EITNV-----KSALTKYKDTQNTADETQDKFHAHVKNANEVEFVGKNGA 1258
QY 640 AFNALDDNDG-----DEQFDNKVVVKTHNSVALADGEHFIVDPDKLS 682
Db 1259 TVSAKTDNNGKHTVTDVAEAKVGDGLEKDTGKIKLVDN-----TDGNNLLT----- 1307
QY 683 STIYDLEKILKXLEGLDYSLRQTTQTOFRDMTTAGKALLSKN-----LRQK 730
Db 1308 ---VDATKGSVAGFEFNAVTTDTAQTGTANERGVVVGSGNGATATETDKKKVATVG 1364
QY 731 EKQFLOBAQFPLGRVDLDKAIKAKEKALVTKKATKNGQLLERSINKAVLAYNNSAIKKA 790
Db 1365 DVAKAINDAATFV-KVEND-----DSATIDDSPTDDG-----ANDALKAXDTLTAKG 1411
QY 791 NVKALEKELDLTLGLVEGKGLAQMVMQGVYLLKPLPEYIYGLNVYFDKSGKL-IY 849
Db 1412 ---KMLKVKRD-----GKNITPALANDLSVK 1434
QY 850 ALDMSDTIGEGQDAYGNPILNVDEDEG-----YHALAVA-TLADYEGLD 894
Db 1435 SATVSDKLSLGTN-----GNKV-NITSDTKGLNPAKSKTGDDANHLNGIASTLTD----- 1485
QY 895 IKTILNS-----KLSQTSIROVPTAAHYHRAGIFQAIQNAAAEAQELLPKP 940
Db 1486 --TLNLSGATNLGNGIGITDNEKKAASVKDVLNAGWVRGVKPPASANNQVENIDFVATY 1543
QY 941 GT-A--HSEKSSSESANSKDRGLQSNPK 965
Db 1544 DTVAFVSGDKDTSVTVESKDKGKRETEVK 1572
```

## RESULT 95

```
US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
```

```
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RPT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-685-467-4

Query Match 3.1%; Score 158.5; DB 3; Length 1912;
Best Local Similarity 18.0%; Pred. No. 0.045;
Matches 211; Conservative 158; Mismatches 388; Indels 412; Gaps 55;

QY 47 YAESGSKLKNITSGPVDVDTVD-----LPSDKRTTPE-----KIKDN 86
Db 566 HVEDAYGELLNLEKNANKQPLVTDSTAATVGDRLKLGWVVSTKNGTKESNQVKQADEV 625
QY 87 LAKGPREQELKAVTEN-----TESEKOITSGSQLEQSKESLSLNTKTPVSTSNWEICDPIT 141
Db 626 LFTGAGAAVTSKSENGKHTITVSAETKADCGLEKGGDTIKLVKVDNQTUN-----VLT 680
QY 142 KGNTLVGLSGSVR-KLSQTDH-----LVLPSSAAD-----GTOLIQVASPAF 183
Db 681 VGNGTAVTKGPFVTKTGATDADRGKVTVKDATTANDADKKVATVKDVATAINSAAATPVK 740
QY 184 TPDKKTAIAE-----YTSRAGENGELISQLDVVGK-----BIINEGEVFN 222
Db 741 TENLTTSIDEDNPDTNGKDDALKAGDTLTTPKAGKNLKVKR---DGNKITFDLAKNLEVKT 797
QY 223 SYLLKKVTI---PTGYKHIGQDAFVDNKNIAEYNLPESLETISDYAPAHALAQIDLPD 278
Db 798 AKVSDTUTIGNTPTG---GTTA-----TPKNITSTADGLN---FA-----KE 835
QY 279 NLKAIGELAFPDNOITGKLSLPQRLMRLAEAPKSNHIK-----SAGAKSSHVDLNVDATKKSNAASIEDVLRAG 888
Db 836 TADASGSKNVYLKGIATTLTEP-----SAGAKSSHVDLNVDATKKSNAASIEDVLRAG 888
QY 318 -TIEFRGNLSKVIG---EASFQDND-----LSQLMLPDG-----LEKIESEA 355
Db 889 WNIQNGNNDYVATYDTVNFDTDDSTGTTTVTVTKADGKADVKIGAKTSVIKDHNGKL 948
QY 356 FTGNPGDDHYNNRVVLWTKSGKNP-SGLATENTVVP-DKSLWQESPEIDYTKWLEEDFT 413
Db 949 FTGKDLKD-ANNGATVSEDDGKTGTGLVTKTVIDAVNKGSGWRVTGE-----GA 997
QY 414 YQKNSVTGFSNKGKQKVRKNKLEIPKQHNGVTITEIGDAPFRVNDPQNKTLRYKDYLE-- 471
Db 998 TAETGATAVNAAGNAETVTSVGNP-KNGNATTATVSKDNGNIN--KYDVNVG 1048
QY 472 -----EVLKPSITIRKIG-----AFAPQSNLKSFEASD 499
Db 1049 DGLKIGDDKKIVADTTTLTVTGKVSVPAGANSVNNKKLVNAEGLATALLNL-SWTAKA 1107
QY 500 DL-----BEIKEG-----AFMNNRIE-----TLELKDKL-----VTIGDAAF 531
Db 1108 DKYADGESEGETDQEVKAGDKVTFKAGKKNLVKQSEKDFTVSLQDTLTGLTSITLGGTAN 1167
QY 532 HINHIYAIVLPESVQ-EIGRSAPFRQNGANNLIPMGSKVKTIGEMAFLSNRLEHLDLSEQK 590
Db 1168 GRNDTGTVINKDGLTITILANGAAAGTDASN-----GNTISVTKDGISAGN-----K 1213
```

```

Qy 591 QLTETPQAFSDNALKEVLLPASLKTTRBEAFKQNHKLQLEVASALSH-----I 639
Db 1214 EITNV-----KSA LKTYKDTQNTADETQDKBEFHAAVKNAVEFVGKNGA 1258
Qy 640 AFNALDDNDG-----DEQFDNKVVVKTHHNSYALADGBHFIVDPDKLS 682
Db 1259 TVSAKTDNNGKHTVTIDVAEAKVGDGLEKOTDGKIKLVKN-----TDGNNLLT----- 1307
Qy 683 STIVDLKILKLEGLDYSLRQTQTQPRDMMTAGKALLSKN-----LRQG 730
Db 1308 ---VDATKGASVAKGEFNAYTTDATTAQGTNANERGVVVKSGNGCATATETDKKCVATVG 1364
Qy 731 EKOKFLQEAOFFLGRVLDLKAIAKAEKALVTKKATKNGQLLERSINKAVLAYNNSAIKKA 790
Db 1365 DVAKAINDAATFV-KVEND-----DSATIDDSPTDGG-----ANDALKAXOTLILKAG 1411
Qy 791 NVRLKLEKLDLLTGLVEKGKGLAQATMVQGYLLKTLPLPEYIYGLNVYFDKSGKI-IY 849
Db 1412 --KNLKVKRD-----GKNITFALANDLSVK 1434
Qy 850 ALDMSDTIGSGKDAVGNPILNVDENEG-----YHALAVA-TLADYEGLD 894
Db 1435 SATVSDKLSLGTN---GNKV-NITSDTKGLNFAKDSKTGDDANIHLNGIATLTD----- 1485
Qy 895 IKTILNS-----KLSQLTSIRQVPTAAHYHRAGIFQAIQNAAAAEQLLPKP 940
Db 1486 --TILNSGATNLLGGNGITDNEKKKAASVKDVLNAGMNVGVKVPASANNQVENIDFVATY 1543
Qy 941 GT----HSEKSSSESANSKDRGLQSNPK 965
Db 1544 DTVDVSGDKDTSVTVESKDNQGRTEVK 1572

RESULT 96
US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-33

```

Qy	184	TPDKKTAIAE-----YTSRAGENGEBISQLDVGDK-----EIIINEGRVFN	222
Db	742	TENLTTSIDBDNPDTDNGKDDALKAGDGLTLTFKAGKNLKVGR---DGNKITFDLAKNLEVK	798
Qy	223	SYLLKKVTI---PTGYKHIGQDAFVDNKNIAEVLNPESLETISDYAFAHALAKLQIDLPD	278
Db	799	AKVSDTLTIGNTPTG---GTTA-----TPKVNITSTADGLN---FA-----KE	836
Qy	279	NLKAIGELAFDFDQITQKLSPLRQLMRAERAFKSNHIK-----SAGAKSSHVDLNV	317
Db	837	TADASGKNVYLKGIATTLTEP-----SAGAKSSHVDLNV	889
Qy	318	TTIEFRGNSLKVIG---EASFQDND---LSQLMLPDG-----LEKIESEA	355
Db	890	WNIQGNNGNVYVATYDVTNFTDSTGTTTVTYQKADGKADGVKIGAKTSVIKDHNGKL	949
Qy	356	FTGNPGDDHNNRVLWTKSGKNP-SGLATENTVWNP-DKSLWQESPEIDYTKWLEBDF	413
Db	950	FTGKDLKD-ANNGATVSEDDGKTGTGLVTAKTVIDAVNKSCHRVTE-----GA	998
Qy	414	YQKNSVTGFSNKGLOKQVRKNKLEIPQHNGVITIETIGDNAFRVDPQNTKLRYKYLE--	471
Db	999	TAETGATVAGNAETAETVTSCTSVNF-KVGNATTATVSKDNGNINV-----KYDVNVG	1049
Qy	472	-----EVLKPSITRKIG-----AFAFQSNLKSFEASD	499
Db	1050	DGLKIGDDKIVADTTTLTVTGKSVSPAGANSNNNKLVLNAGEGLATALNNL-SWTAKA	1108
Qy	500	DL-----EEIKEG-----AFMNNRIE-----TLELKDKL-----VTIGDAAF	531
Db	1109	DKYADGESEGETDOEVKAGDKVTFPKAGKNLKVQSEKDFYISLQDTLTGLTSITLGGTAN	1168
Qy	532	HIMHIIAIVLPESVQ-EIGRSAPRQNGANNLI FMGSKVITLGEMAFUSNREHLHDLSEQK	590
Db	1169	GRNDTGTVINKGDLTITLANGAAAGTDASN---GNTISVTKDGISAGN-----K	1214
Qy	591	QLTEIIPQAFSDNALKEVLLPASIKTIREBAFKKNHLKQLEVASALSH-----I	639
Db	1215	EITNV-----KSALKYKTQNTADETQDKFEHAAVKNANEVEFVGKNGA	1259
Qy	640	AFNALDDNDG-----DEQFNKVVVYKTHHNSYALADGHEFIVDPDKLS	682
Db	1260	TVSAKTDDNNGKHTVTIDVAEAKVGDGLEKDTGDKIKLVQDN---TDGNLLT-----	1308
Qy	683	STIVLEKILKLTGDLDYSLRQTTQTFQDMTTAGKALLSKSN-----LRQ	730
Db	1309	---VDATKGSVAKEGFNAVTTDATTAQGTNANERGVVVKSGNGATATETDKKKVATVG	1365
Qy	731	EKQFLQEAQFFLGRVOLDKAIKAEBKALVTKATKNGQLLERSINKAVLAYNNSAIKKA	790
Db	1366	DVAKAINDAATFV-KVEND-----DSATIDDSPTDDGA-----NDALKAG	1404
Qy	791	NVNRLEKELDLLTGLVEGKGPLAQTWVGQVYLLTKPLPLPYEYIIGLVNVPYDKSGKL-IY	849
Db	1405	-----DLTL-----LKAGKNLKVKRDGKNITFALANDLSVK	1435
Qy	850	ALDMSDTIGEGQKDAYCNPIILNVDEDNeg-----YHALAVA-TLADYEGLD	894
Db	1436	SATVSKLSLGTN---GNKV-NITSDTKGLNPAKDKTGDANDIHLNGIASTLTD-----	1486
Qy	895	IKTILNS-----KLSQLTISRQVPTAAYHRAIGIFQATIQNAAAAEQLPKP	940
Db	1487	--TLANSGATTNLGGNGITDNEKKRAASVKDVLNAGNVGVKVPASANNQVENIDFVATY	1544
Qy	941	GT-----HSEKSSSESSEANSKDRGLQSNPK	965
Db	1545	DTYDFVSGDKDTSVTTVESKDNKGRTEVK	1573





```
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-669-974-33

Query Match          3.1%; Score 158.5; DB 3; Length 2353;
Best Local Similarity 18.2%; Pred. No. 0.062;
Matches 213; Conservative 155; Mismatches 389; Indels 412; Gaps 55;

QY 47 YAESSGSKLKNINETSQGVDDTVTD-----LFSDKRTTPE-----KIKDN 86
DB 567 HVEDAYKGLNLNKNANKQPLVTDSTAATVGDRLKLGWVSTKNGTKEESNQVQADEV 626
QY 87 LAKPREQELKAVTEN-----TSEKQITSGSOLEKESLSLNTKVTSTSNWEICDFIT 141
DB 627 LFTGAGAATVTSKSENGKHTITVSAETKADCGLEKGDGTIKLVKDNQNTDN-----VLT 681
QY 142 KGNLTVGLSKSGVE--KLSQTDH-----LVLPQAAD-----GTQLIQVASPAF 183
DB 682 VGNNGTAVTKGPFETVKGTATDADRKGVTVKDANDADKKVATVKDVATAINSAATFVK 741
QY 184 TPKKKTAAIE-----YTSRAGEGEISQLDVDGK-----EINEGEVFN 222
DB 742 TENLTTSIDEDNPTDNGKDDALKAGDTLTFKAGKNLKVVR---DGKNITFDLAKNLEVK 798
QY 223 SYLLKKVTI-----PTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAPHAHLAKQIDLP 278
DB 799 AKVSDTLTIGGNTPTG---GTTA-----TPKVNITSTADGLN---FA-----KE 836
QY 279 NLKAIGELAFEDNQITGKLSLPRQLMRLAERAFKSNHIK-----KIDN 317
DB 837 TADASGKNVYLKGIATTLTEP-----SAGAKSSHVDLNVNDAATKKSNAASIEDVLRAG 889
QY 318 -TIEFRGNSLKVIG---EASFQDND-----LSQLMLPDG-----LEKIESEA 355

RESULT 99
US-09-797-862-33
; Sequence 33, Application US/09797862
; Patent No. 6607729
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT FILING DATE: 2001-05-03
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-797-862-33

Query Match          3.1%; Score 158.5; DB 4; Length 2353;
Best Local Similarity 18.2%; Pred. No. 0.062;
Matches 213; Conservative 155; Mismatches 389; Indels 412; Gaps 55;

QY 47 YAESSGSKLKNINETSQGVDDTVTD-----LFSDKRTTPE-----KIKDN 86
DB 567 HVEDAYKGLNLNKNANKQPLVTDSTAATVGDRLKLGWVSTKNGTKEESNQVQADEV 626
QY 87 LAKPREQELKAVTEN-----TSEKQITSGSOLEKESLSLNTKVTSTSNWEICDFIT 141
DB 627 LFTGAGAATVTSKSENGKHTITVSAETKADCGLEKGDGTIKLVKDNQNTDN-----VLT 681
QY 142 KGNLTVGLSKSGVE--KLSQTDH-----LVLPQAAD-----GTQLIQVASPAF 183
DB 682 VGNNGTAVTKGPFETVKGTATDADRKGVTVKDANDADKKVATVKDVATAINSAATFVK 741
QY 184 TPKKKTAAIE-----YTSRAGEGEISQLDVDGK-----EINEGEVFN 222
DB 742 TENLTTSIDEDNPTDNGKDDALKAGDTLTFKAGKNLKVVR---DGKNITFDLAKNLEVK 798
QY 223 SYLLKKVTI-----PTGYKHIGQDAFVDNKNIAEVLNLPESLETISDYAPHAHLAKQIDLP 278
DB 799 AKVSDTLTIGGNTPTG---GTTA-----TPKVNITSTADGLN---FA-----KE 836
QY 279 NLKAIGELAFEDNQITGKLSLPRQLMRLAERAFKSNHIK-----KIDN 317
DB 837 TADASGKNVYLKGIATTLTEP-----SAGAKSSHVDLNVNDAATKKSNAASIEDVLRAG 889
QY 318 -TIEFRGNSLKVIG---EASFQDND-----LSQLMLPDG-----LEKIESEA 355
```

```
Db      890 WNIQGNVNDVAVYDVTNFTDDSTGTTVTVTQKADGKADGVKIGAKTSVIKDHNGKL 949
Qy      356 FTGNGPDHNNRVVLTWTKSGKNP-SGLATENTYVNP-DKSLWQSSPEIDYTKWLEEDFT 413
Db      950 FTGDKLKD-ANNGATVSEDDGKDTGTGLVTAKTVIDAVNKSQWRTGE-----GA 998
Qy      414 YQKNSVTGFSNKGLOKVRNKNLETPKQHNGVTTIETIGDNAPRVDFQNKTLRKYDLE-- 471
Db      999 TAETGATAVNAGNAETVTSVNF-KNGNATTATVSKDNGNINV-----KYDVNVG 1049
Qy      472 -----EVLKPSITIRKIG-----AFAFQSNLNKSFASD 499
Db      1050 DLKIGDDKKIVADTTTLVTGCKVSVFAGANSVNNKLVNAEGLATLNNL-SWTAKA 1108
Qy      500 DL-----EBIKEG-----AFMNRIE-----TLELKKL-----VTIGDAAF 531
Db      1109 DKYADGSEGETDQSVKAGDKVTTFKAGKNLVKQKSEKDTFTYSLQDTLGLTSITLGGTAN 1168
Qy      532 HINHIAIVLPESVQ-EIGRSAPFRONGANNLIFMGSKVKTIGEMAFPLNRLHLDLSQK 590
Db      1169 GRNDTGTWINKDGLTITLANGAAAGTASN-----GNTISVTKDGISAGN-----K 1214
Qy      591 OLTEIPVQAFSDNALKEVLLPASLTIREEAFKKHKLQLEVASALSH-----I 639
Db      1215 EITNV-----KSALETYKDTQNTADETQKEFHAAVKNANEVEFVKGKGA 1259
Qy      640 AFNALDDNDG-----DEQFNKVVVVKTHNSYALADGEHPIDVDPKLS 682
Db      1260 TVSAKTNNNGKHTVTDVAEAKVGDGLEDKTDGKIKLVKN-----TDGNLLT----- 1308
Qy      683 STIVOLEKILKLEGLDVTSLRQTQTQTPRDMTTAGKALLSKN-----LRQK 730
Db      1309 ---VDATKGASVAKGEFAVTTTATTAQGTNANERGVVVGSGNGATATETDKKVVATVG 1365
Qy      731 EKQFLQBAQFPLGRVDLDKALAKAEKALVTKATKNGQLLERSINKAVLAVNNSAIKKA 790
Db      1366 DVAKAINDAATV-KVEND-----DSATIDDSPTDDG-----NDALKAG 1404
Qy      791 NVKLEKELDLTLGLVEGKGPLAQATWQGVVLLKPLPLPEYIYGLNVYFDKSGKL-IY 849
Db      1405 -----DTLT-----LKAGKNLVKVRDGNKITFALANDLSVK 1435
Qy      850 ALMDSMTTGEQKDAYGNPILNVDENEG-----YHALAVA-TLADYEGLD 894
Db      1436 SATVSDKLSLGTN---GNKV-NITSDTKGLNPAKDSKTGDDANIHLNGIASTLTD----- 1486
Qy      895 IKTILNS-----KLSQTSIRQVPTAAVHRAGIFQAIQNAAEAEQLPKP 940
Db      1487 ---LLNSGATNVLGNGITDNEKKRAASVKDVLNAGNVRGVKPPASANNQVENIDFVATY 1544
Qy      941 GT-----HSEKSSSESANSKDRGLQSNPK 965
Db      1545 DTVDVSGDKDTSVTVESKONGKTEVK 1573
```

## RESULT 100

US-09-684-707-4

; Sequence 4, Application US/09684707

; Patent No. 6759213

; GENERAL INFORMATION:

; APPLICANT: St. Gene, Joseph

; Barenkamp, Stephen J.

; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESS: Flehr Hobbach Test Albritton &amp; Herbert LLP

; STREET: Four Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: United States

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

```
;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/684,707
; FILING DATE: 04-Oct-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; APPLICATION NUMBER: PC7/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/RT/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2353 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-684-707-4
```

Query Match 3.1%; Score 158.5; DB 4; Length 2353;

Best Local Similarity 18.2%; Pred. No. 0.062;

Matches 213; Conservative 155; Mismatches 389; Indels 412; Gaps 55;

```
Qy      47 YAESGSGSKLINETSQVDDTVD-----TESEKQITSGSQLEQKESLSLNKTVSPSTNWEICDPT 141
Db      567 HVEDAYKGLNLNENKANKQPLVTDSTAATVGDRLKLGWVVSTKNGTKTEENQVQKQAD 626
Qy      87 LAKGPREQELKAVTEN-----TESEKQITSGSQLEQKESLSLNKTVSPSTNWEICDPT 141
Db      627 LFTGAGAAVTVSKSENGKHITVSAETKADCGLEKDGDTIKLVKVDNQNTDN-----VLT 681
Qy      142 KGNTLVLSKSGVE--KLSQTDH-----LVLPSSAAD-----GTOLIQVASPAF 183
Db      682 VGNNGTAVTKGPFETVKGTADADRGKVTVDATANDADKKVATVKDVAATAINSAAATPVK 741
Qy      184 TPDKKTAAIE-----YTSBAGENGESIQLDVDGK-----EINIEGEVFN 222
Db      742 TENLTTSIDEDNPTDNGKODALKAGDTLTFKAGKNLVKVR---DGKNITFDLAKNLEVK 798
Qy      223 SYLLKVKVTI---PTGYKHIGQDAPVDNKNIAEYNLPESLETISDYAFALHAKQLDLPD 278
Db      799 AKVSDTLTIGGNTPTG---GTTA-----TPKNITSTADGLN---FA-----KE 836
Qy      279 NLKAIAGLAPFDNQITKLSLPLQMLRAELAPKSNHIK-----SAGAKSSHVDLNVDA 317
Db      837 TADAGSGKNVYLKGIATTLTEP-----SAGAKSSHVDLNVDAKKSNAAASIEDVLRAG 889
Qy      318 -TIEFRGNLSKVIG--EASFQDND-----LSQLMLPDG-----LEKISEA 355
Db      890 WNIQGNVNDVAVYDVTNFTDDSTGTTVTVTQKADGKADGVKIGAKTSVIKDHNGKL 949
Qy      356 FTGNGPDHNNRVVLTWTKSGKNP-SGLATENTYVNP-DKSLWQSSPEIDYTKWLEEDFT 413
Db      950 FTGDKLKD-ANNGATVSEDDGKDTGTGLVTAKTVIDAVNKSQWRTGE-----GA 998
Qy      414 YQKNSVTGFSNKGLOKVRNKNLEIPKQHNGVTTIETIGDNAPRVDFQNKTLRKYDLE-- 471
Db      999 TAETGATAVNAGNAETVTSVNF-KNGNATTATVSKDNGNINV-----KYDVNVG 1049
Qy      472 -----EVLKPSITIRKIG-----AFAFQSNLNKSFASD 499
```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:50:07 ; Search time 182 Seconds  
(without alignments)  
2836.131 Million cell updates/sec

Title: US-10-078-531-2  
Perfect score: 1008  
Sequence: 1 MKHLKTVALLTTVSWVTH.....LGYTSVALLTAIKKKY 1008

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 110 summaries

Database : Uniprot\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1008	100.0	1008	2	Q9A0C0 streptococ
2	679	67.4	1008	2	Q8P1F7 streptococ
3	293	29.1	1008	2	Q878I7 streptococ
4	283	28.1	999	2	Q8K7X8 streptococ
5	89	8.8	1055	2	Q8S6V3 streptococ
6	76	7.5	1055	2	Q8E1D6 streptococ
7	9	0.9	245	2	Q9ZRT5 streptococ
8	9	0.9	279	2	Q9NW78 homo sapien
9	9	0.9	455	2	Q86A18 dictyosteli
10	9	0.9	523	2	Q8C134 mus musculu
11	9	0.9	564	2	Q80TA6 mus musculu
12	9	0.9	568	2	Q9D2Y7 mus musculu
13	9	0.9	583	2	Q86X27 homo sapien
14	9	0.9	590	2	Q9FHE1 arabidopsis
15	9	0.9	590	2	Q7TPZ6 mus musculu
16	9	0.9	590	2	Q8BZU2 mus musculu
17	9	0.9	590	2	Q9ERD6 mus musculu
18	9	0.9	907	1	LGR5_HUMAN
19	8	0.8	12	2	Q75MW9 homo sapien
20	8	0.8	67	2	Q982N8 rhizobium
21	8	0.8	67	2	Q7UVJ9 rhodospirell
22	8	0.8	106	1	RL3E_SULTO
23	8	0.8	116	2	Q9NSZ8 homo sapien
24	8	0.8	165	2	Q87WF0 pseudomonas
25	8	0.8	174	2	Q9SB23 hordeum vul
26	8	0.8	237	2	Q9FR02 nicotiana t
27	8	0.8	240	2	Q8TW63 methanopyru
28	8	0.8	241	2	Q7T3A2 brachydanio
29	8	0.8	242	2	Q7NST9 chromobacte
30	8	0.8	253	2	P93664 chlamydomon
31	8	0.8	255	2	Q8AB78 bacteroides

Q8UYE5	255	2	Q8UYE5	8	0.8
Q6YLZ5	255	2	Q6YLZ5	8	0.8
Q776C0	255	2	Q776C0	8	0.8
Q9H0J8	268	2	Q9H0J8	8	0.8
Q96191	284	2	Q96191	8	0.8
Q926G6	296	2	Q926G6	8	0.8
Q9ABZ6	304	2	Q9ABZ6	8	0.8
Q88U35	316	2	Q88U35	8	0.8
YDZA_SCHPO	329	1	YDZA_SCHPO	8	0.8
Q9927	348	2	Q9927	8	0.8
Q99944	348	2	Q99944	8	0.8
Q8M518	348	2	Q8M518	8	0.8
Q8HLA4	348	2	Q8HLA4	8	0.8
Q26490	351	2	Q26490	8	0.8
Q8RPN6	360	2	Q8RPN6	8	0.8
Q9F7S2	360	2	Q9F7S2	8	0.8
Q87880	370	2	Q87880	8	0.8
Q7QF26	385	2	Q7QF26	8	0.8
Q63C90	391	2	Q63C90	8	0.8
Q739H8	391	2	Q739H8	8	0.8
Q81RG8	391	2	Q81RG8	8	0.8
Q6HJ01	391	2	Q6HJ01	8	0.8
ACK2_RHIME	393	1	ACK2_RHIME	8	0.8
Q6BJ95	416	2	Q6BJ95	8	0.8
ARO8_STRMU	427	1	ARO8_STRMU	8	0.8
Q8CCK5	427	2	Q8CCK5	8	0.8
Q97KA1	429	2	Q97KA1	8	0.8
Q9NKW1	441	2	Q9NKW1	8	0.8
ARDY_BACHD	458	1	ARDY_BACHD	8	0.8
Q7VVB6	463	2	Q7VVB6	8	0.8
Q7WKU2	463	2	Q7WKU2	8	0.8
Q9VE79	478	2	Q9VE79	8	0.8
Q9NW94	483	2	Q9NW94	8	0.8
Q6LUN0	485	2	Q6LUN0	8	0.8
Q42989	488	2	Q42989	8	0.8
Q42990	488	2	Q42990	8	0.8
Q6Z5B2	488	2	Q6Z5B2	8	0.8
Q8WPP2	490	2	Q8WPP2	8	0.8
Q6C9U7	492	2	Q6C9U7	8	0.8
AMYB_WHEAT	503	1	AMYB_WHEAT	8	0.8
GPWI_MYCGA	503	1	GPWI_MYCGA	8	0.8
Q8B335	517	2	Q8B335	8	0.8
Q6SNP7	517	2	Q6SNP7	8	0.8
Q7MD66	528	2	Q7MD66	8	0.8
Q9FUK6	533	2	Q9FUK6	8	0.8
Q9FUK7	533	2	Q9FUK7	8	0.8
AMYB_HORVU	535	1	AMYB_HORVU	8	0.8
P82993	535	2	P82993	8	0.8
Q84T19	535	2	Q84T19	8	0.8
Q84T20	535	2	Q84T20	8	0.8
Q9AVJ8	535	2	Q9AVJ8	8	0.8
Q9FSI3	535	2	Q9FSI3	8	0.8
Q9SBH7	535	2	Q9SBH7	8	0.8
Q7X3X6	543	2	Q7X3X6	8	0.8
Q6V1X2	546	2	Q6V1X2	8	0.8
Q6WKW9	554	2	Q6WKW9	8	0.8
Q38914	555	2	Q38914	8	0.8
Q42462	555	2	Q42462	8	0.8
Q9P4A3	594	2	Q9P4A3	8	0.8
Q7VQZ4	625	2	Q7VQZ4	8	0.8
Q9FFV8	642	2	Q9FFV8	8	0.8
Q6ZPU7	649	2	Q6ZPU7	8	0.8
Q6MW74	662	2	Q6MW74	8	0.8
Q7XP38	673	2	Q7XP38	8	0.8
Q7Z5E6	681	2	Q7Z5E6	8	0.8
Q54781	681	2	Q54781	8	0.8
Q8VCD9	682	2	Q8VCD9	8	0.8
P78362	686	2	P78362	8	0.8
Q81YQ3	688	2	Q81YQ3	8	0.8
Q6NULO	688	2	Q6NULO	8	0.8
Q7NRC2	719	2	Q7NRC2	8	0.8
Q98PD0	757	2	Q98PD0	8	0.8



```
105      8      0.8      776      2      Q9AAE9      Q9aae9 caulobacter
106      8      0.8      784      1      TLR2_MAFCA      Q95m53 macaca fasc
107      8      0.8      819      2      Q8CIG4      Q8cig4 mus musculu
108      8      0.8      862      2      Q9P5X6      Q9p5x6 neurospora
109      8      0.8      879      2      O76978      O76978 scypha raph
110      8      0.8      880      1      SYV_BACST      P11931 bacillus st

RESULT 1
Q9AAC0
ID      Q9AAC0      PRELIMINARY;      PRT; 1008 AA.
AC      Q9AAC0;
DT      01-JUN-2001 (TReMBLrel. 17, Created)
DT      01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT      01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE      Hypothetical protein SPY0843.
GN      OrderedLocustNames=SPY0843;
OS      Streptococcus pyogenes.
OC      Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC      Streptococcus.
OX      NCBI_TaxID=1314;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX      MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071593998;
RA      Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Lyon K.,
RA      Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA      Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA      Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
RA      "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RT      Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
RL      EMBL; AE006534; AAK33772.1; -.
DR      GO; GO:000986; C:cell surface; ISA.
DR      InterPro; IPR001899; Gram_pos_anchor.
DR      InterPro; IPR007093; LRR_Tp.
DR      PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW      Complete proteome; Hypothetical protein.
SQ      SEQUENCE 1008 AA; 111503 MW; 045BP8CE931AF0CF CRC64;
```

```
Query Match      100.0%; Score 1008; DB 2; Length 1008;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      MKKLKLTVALTITTVSVTHNQEVSLVKEPIKQTOASSISGADYAESGSKLKINE 60
DB      1      MKKLKLTVALTITTVSVTHNQEVSLVKEPIKQTOASSISGADYAESGSKLKINE 60

QY      61      TSGPVDVDTVDLFSDKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120
DB      61      TSGPVDVDTVDLFSDKRTTPEKIKONLAKPREQELKAVTENSEKQITSGSQLEQSK 120

QY      121      SISLNTKVPSTNWEICDITKGNLTGLVLSKGVKLSQTDHLVPSQAADGTQLIQVAS 180
DB      121      SISLNTKVPSTNWEICDITKGNLTGLVLSKGVKLSQTDHLVPSQAADGTQLIQVAS 180

QY      181      FAFTPDKKTAAIAYTSRAGENGESIQDVGKEIINEGEVFNYSLLKKVTIPTGYKHGQ 240
DB      181      FAFTPDKKTAAIAYTSRAGENGESIQDVGKEIINEGEVFNYSLLKKVTIPTGYKHGQ 240

QY      241      DAFVONKNAIENVLPESLETISDYAPAHALKQIDLPONLKAIGBLAFPDNQITKLSLP 300
DB      241      DAFVONKNAIENVLPESLETISDYAPAHALKQIDLPONLKAIGBLAFPDNQITKLSLP 300

QY      301      ROLMLAERAFKSNHKTTEPGNSIKVIGESFQNDLSQMLPDGLEKIESEFTGNP 360
DB      301      ROLMLAERAFKSNHKTTEPGNSIKVIGESFQNDLSQMLPDGLEKIESEFTGNP 360

QY      361      GDHYNNRVVLWTKGKQPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB      361      GDHYNNRVVLWTKGKQPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
```

```
QY      421      GFSNKGLOKVRKNLEIPKQHNGVTITIGDNAFRNVDPONKTLRKVDLEEVKLPSTIR 480
DB      421      GFSNKGLOKVRKNLEIPKQHNGVTITIGDNAFRNVDPONKTLRKVDLEEVKLPSTIR 480

QY      481      KIGAFAPQSNLKGFEASDDLEIEKEGAFNMNRITETLEKDKLVITIGDAAAFHINHIYAIV 540
DB      481      KIGAFAPQSNLKGFEASDDLEIEKEGAFNMNRITETLEKDKLVITIGDAAAFHINHIYAIV 540

QY      541      LPESVQEIGRSAFRQNGANNLIIPGSKVKTIGEMAFISNRLEHLDLSEKQKLTIPVQAF 600
DB      541      LPESVQEIGRSAFRQNGANNLIIPGSKVKTIGEMAFISNRLEHLDLSEKQKLTIPVQAF 600

QY      601      SDNALKEVLLPASLKTITREAFKKNHLKQLEVASALSHIAFNALDDNDGDQFQNKVVVK 660
DB      601      SDNALKEVLLPASLKTITREAFKKNHLKQLEVASALSHIAFNALDDNDGDQFQNKVVVK 660

QY      661      THNSYALADGEHFIVDPDKLSSTIVDEKILKIEGLDYSTLRQTTQTQFRDMTTAGKA 720
DB      661      THNSYALADGEHFIVDPDKLSSTIVDEKILKIEGLDYSTLRQTTQTQFRDMTTAGKA 720

QY      721      LLSKSNLRQGEKQFLQEAQFPLGRVLDLKAIAKAERKALVTKATKQGLLERSINKAVL 780
DB      721      LLSKSNLRQGEKQFLQEAQFPLGRVLDLKAIAKAERKALVTKATKQGLLERSINKAVL 780

QY      781      AYNSAIKKNVKNLEKELDLITGLVSGKGLAQATMVQGVYLLKTLPLPEYYIGLVNY 840
DB      781      AYNSAIKKNVKNLEKELDLITGLVSGKGLAQATMVQGVYLLKTLPLPEYYIGLVNY 840

QY      841      FDKSGKLIYALDMSDTTIGEGQDAYGNPILNVDDNDEGYHALAVATLADYEGLDIKTILN 900
DB      841      FDKSGKLIYALDMSDTTIGEGQDAYGNPILNVDDNDEGYHALAVATLADYEGLDIKTILN 900

QY      901      SKLSQLTISIRQVPTAAVHRAGIFQAIQNAAEAEQLLPKGTSEKSSSSSANSKDRGL 960
DB      901      SKLSQLTISIRQVPTAAVHRAGIFQAIQNAAEAEQLLPKGTSEKSSSSSANSKDRGL 960

QY      961      QSNPKTNRGRHSAILPRTGSKGSFVYGLGTVTSVALLSLITAIKKKKY 1008
DB      961      QSNPKTNRGRHSAILPRTGSKGSFVYGLGTVTSVALLSLITAIKKKKY 1008

RESULT 2
Q8P1F7
ID      Q8P1F7      PRELIMINARY;      PRT; 1008 AA.
AC      Q8P1F7;
DT      01-OCT-2002 (TReMBLrel. 22, Created)
DT      01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT      01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE      Hypothetical protein spyM18_0903.
GN      OrderedLocustNames=spyM18_0903;
OS      Streptococcus pyogenes (serotype M18).
OC      Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC      Streptococcus.
OX      NCBI_TaxID=186103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MGAS8232;
RX      MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;
RA      Smoot J.C., Barbhan K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA      Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA      Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA      Kapur V., Daly J.A., Veasy L.G., Mueser J.M.;
RA      "Genome sequence and comparative microarray analysis of serotype M18
RT      group A Streptococcus strains associated with acute rheumatic fever
RT      outbreaks.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR      EMBL; AE010019; AAL9753.1; -.
DR      GO; GO:000986; C:cell surface; IEA.
DR      InterPro; IPR001899; Gram_pos_anchor.
DR      InterPro; IPR007093; LRR_Tp.
DR      PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW      Complete proteome; Hypothetical protein.
```

RESULT 3	
Q878I7	
ID Q878I7	PRELIMINARY; PRT; 1008 AA.
AC Q878I7;	
DT 01-JUN-2003	(TREMBLrel. 24, Created)
DT 01-JUN-2003	(TREMBLrel. 24, Last sequence update)

Db 541 LPESVQIGRSFRQNGANNLIFMGSKVKTIGEMAFNSLRLEHLDLSEKQLTETPVAQF 600  
Qy 601 SDNALKEVLLPASLTKTIRREAFKQHLKQLEVASALSALSHIAFNALDDNDGDFQFNKVYK 660  
Db 601 SDNALKEVLLPASLTKTIRREAFKQHLKQLEVASALSALSHIAFNALDDNDGDFQFNKVYK 660  
Qy 661 THHNSYALADGHRHFDVDPDKLSTIVDLKILKLEGLDYSLRTOTQPRDMTTAGKA 720  
Db 661 THHNSYALADGHRHFDVDPDKLSTIVDLKILKLEGLDYSLRTOTQPRDMTTAGKA 720  
Qy 721 LLSKSNLRQGEKQFLQEAQFFLGRVDLDKATAKAELKALVTKATKNGQLLERSINKAVL 780  
Db 721 LLSKSNLRQGEKQFLQEAQFFLGRVDLDKATAKAELKALVTKATKNGQLLERSINKAVL 780  
Qy 781 AYNSAIIKANKVRLKLELDLTLGLVEGKPLAQATMWQVYLLKTPPLPBYIYGLNVY 840  
Db 781 AYNSAIIKANKVRLKLELDLTLGLVEGKPLAQATMWQVYLLKTPPLPBYIYGLNVY 840  
Qy 841 FPKSGKLIYALDMSDTIGEGQDAYGNPILNVDENEGYSALAVATLADYGLDITILN 900  
Db 841 FPKSGKLIYALDMSDTIGEGQDAYGNPILNVDENEGYSALAVATLADYGLDITILN 900  
Qy 901 SKLSQLTSTRQVPTAAYHRAGIFQAIQNAAAAEQLLPKPGTHSEKSSSESANSKDRGL 960  
Db 901 SKLSQLTSTRQVPTAAYHRAGIFQAIQNAAAAEQLLPKPGTHSEKSSSESANSKDRGL 960  
Qy 961 QSNPKTNRGRHSAIIPRTGSKGSFVYGLGYTSVALLSLI 1000  
Db 961 QSNPKTNRGRHSAIIPRTGSKGSFVYGLGYTSVALLSLI 1000

RESULT 4  
Q8K7X8 ID Q8K7X8 PRELIMINARY; PRT; 999 AA.  
AC Q8K7X8;  
DT 01-OCT-2002 (TREMELrel. 22, Created)  
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)  
DE Putative surface antigen.  
GN OrderedLocusNames=SpvM3\_0569;  
OS Streptococcus pyogenes (serotype M3).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=198466;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MGAS315 / Serotype M3;  
RX MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;  
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,  
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,  
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,  
RA Schlievert P.M., Musser J.M.;  
RT "Genome sequence of a serotype M3 strain of group A Streptococcus;  
RT phage-encoded toxins, the high-virulence phenotype, and clone  
RT emergence.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).  
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
CC an amide bond (By similarity).  
CC EMBL; AS014148; AA079176.1; -;  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR007093; LRR\_Tp.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR TIGRfam; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
KW Cell wall; Complete proteome; Peptidoglycan-anchor.  
SQ SEQUENCE 999 AA; 110518 MW; 39F9801882BAEF7C CRC64;

Query Match 28.1%; Score 283; DB 2; Length 999;  
Best Local Similarity 99.3%; Pred. No. 1e-276; Indels 0; Gaps 0;  
Matches 983; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 11 TLTTVSVTHNQEVFSLVKEPILKQTOASSISGADYAESGSKGLKINETSQVDDTVT 70  
Db 2 TLTTVSVTHNQEVFSLVKEPILKQTOASSISGADYAESGSKGLKINETSQVDDTVT 61  
Qy 71 DLFSDKRTTPEKIKDNIAKGPREQELKAVTENTESEKQITSGSQLEQSKESLSLNTKTPS 130  
Db 62 DLFSDKRTTPEKIKDNIAKGPREQELKAVTENTESEKQITSGSQLEQSKESLSLNTKTPS 121  
Qy 131 TSNWEICDFTKNTGNTLVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVAFPTPKKTA 190  
Db 122 TSNWEICDFTKNTGNTLVGLSKSGVEKLSQTDHLVLPQAAADGTQLIQVAFPTPKKTA 181  
Qy 191 IAEYTSRAGENGEISQLDVKGKEIINEGEVFNPSYLLAKKVTIPTGYKHIGQDAFVNDKNIA 250  
Db 182 IAEYTSRAGENGEISQLDVKGKEIINEGEVFNPSYLLAKKVTIPTGYKHIGQDAFVNDKNIA 241  
Qy 251 EVNLPESLETISDYAFALHALKQIDLPNLKAIAGELAFFDNQITGKLSLPRQLMRLAERA 310  
Db 242 EVNLPESLETISDYAFALHALKQIDLPNLKAIAGELAFFDNQITGKLSLPRQLMRLAERA 301  
Qy 311 FKSNIHITIEPRGNLSKVIGEASPDNDLSQMLPDGLEKIESEAFNPGDDHYNRVV 370  
Db 302 FKSNIHITIEPRGNLSKVIGEASPDNDLSQMLPDGLEKIESEAFNPGDDHYNRVV 361  
Qy 371 LMTSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSKGLQKV 430  
Db 362 LMTSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSKGLQKV 421  
Qy 431 KRKNLEIPKQHNGVTITEIGDNAFRNVDFONKTLRYDLEEVKLPSTIRKIGAFQSN 490  
Db 422 KRKNLEIPKQHNGVTITEIGDNAFRNVDFONKTLRYDLEEVKLPSTIRKIGAFQSN 481  
Qy 491 NLKSFEASDDLEEKEGAFMNNRIETLEKDKLVTIGDAAPHINHIYAIVLPESVQIEGR 550  
Db 482 NLKSFEASDDLEEKEGAFMNNRIETLEKDKLVTIGDAAPHINHIYAIVLPESVQIEGR 541  
Qy 551 SAFQNGANNLI FMGSKVKTIGEMAFNSLRLEHLDLSEKQLTETPVAQFSDNALKEVLL 610  
Db 542 SAFQNGANNLI FMGSKVKTIGEMAFNSLRLEHLDLSEKQLTETPVAQFSDNALKEVLL 601  
Qy 611 PASLKTIREAFKKNHLKQLEVASALSALSHIAFNALDDNDGDFQFNKVYVTHHNSYALAD 670  
Db 602 PASLKTIREAFKKNHLKQLEVASALSALSHIAFNALDDNDGDFQFNKVYVTHHNSYALAD 661  
Qy 671 GEHFIVDPDKLSSTIVLEKILKLEGLDYSLRTOTQPRDMTTAGKALLSKLRQ 730  
Db 662 GEHFIVDPDKLSSTIVLEKILKLEGLDYSLRTOTQPRDMTTAGKALLSKLRQ 721  
Qy 731 EKQFLQEAQFFLGRVDLDKATAKAELKALVTKATKNGQLLERSINKAVLAYNNSAIKKA 790  
Db 722 EKQFLQEAQFFLGRVDLDKATAKAELKALVTKATKNGQLLERSINKAVLAYNNSAIKKA 781  
Qy 791 NVKRLKELDLTLGLVEGKPLAQATMWQVYLLKTPPLPBYIYGLNVYFDKSGKLIYA 850  
Db 782 NVKRLKELDLTLGLVEGKPLAQATMWQVYLLKTPPLPBYIYGLNVYFDKSGKLIYA 841  
Qy 851 LDMSDTIGEGQDAYGNPILNVDENEGYSALAVATLADYGLDITILNSKLSQLTISR 910  
Db 842 LDMSDTIGEGQDAYGNPILNVDENEGYSALAVATLADYGLDITILNSKLSQLTISR 901  
Qy 911 QVPTAAYHRAGIFQAIQNAAAAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGR 970  
Db 902 QVPTAAYHRAGIFQAIQNAAAAEQLLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGR 961  
Qy 971 HSAILPRTGSKGSFVYGLGYTSVALLSLI 1000  
Db 962 HSAILPRTGSKGSFVYGLGYTSVALLSLI 991

RESULT 5  
Q8E6V3 ID Q8E6V3 PRELIMINARY; PRT; 1055 AA.  
AC Q8E6V3;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein gbs0456.  
 GN OrderedLocusNames=gbs0456;  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216495;  
 RN (1) \_\_\_\_\_  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaeser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,  
 RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease.";  
 RL Mol. Microbiol. 45:1499-1513 (2002).  
 DR EMBL; AL766845; CAD46100.1; -;  
 DR Sagalish; gbs0456; -;  
 DR InterPro; IPR007093; LRR\_Tp.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 1055 AA; 117976 MW; 9DD56FBAD171B98 CRC64;  
 Query Match 8.8%; Score 89; DB 2; Length 1055;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-80;  
 Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 (1) \_\_\_\_\_  
 QY 805 LVGEKGLAQAATWVGQVYLLKTPPLPEYIYGLNLYFVFKSGKLIYALDMSDTIGSEQKDA 864  
 DB 806 LVGEKGLAQAATWVGQVYLLKTPPLPEYIYGLNLYFVFKSGKLIYALDMSDTIGSEQKDA 865  
 QY 865 YGNPILNVDEDEGHALAVATLADYEG 893  
 DB 866 YGNPILNVDEDEGHALAVATLADYEG 894  
 RESULT 6  
 Q8E1D6 PRELIMINARY; PRT; 1055 AA.  
 AC Q8E1D6;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Cell wall surface anchor family protein.  
 GN OrderedLocusNames=SAG0421;  
 OS Streptococcus agalactiae (serotype V).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216466;  
 RN (1) \_\_\_\_\_  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2603 V/R / Serotype V;  
 RX MEDLINE=2222998; PubMed=12200547; DOI=10.1073/pnas.182380799;  
 RA Tettelin H., Wasigann V., Cieleszcz M.J., Eisen J.A., Peterson S.N.,  
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,  
 RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,  
 RA DeBoy R.T., Durkin A.S., Kolony J.F., Madupu R., Lewis M.R.,  
 RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,  
 RA Carthy H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,  
 RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,  
 RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative genomic analysis of an  
 RT emerging human pathogen, serotype V Streptococcus agalactiae.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).  
 DR EMBL; AE014211; AAM99327.1; -;  
 DR TIGR; SAG0421; -;  
 DR InterPro; IPR007093; LRR\_Tp.  
 KW Complete proteome.  
 SQ SEQUENCE 1055 AA; 118129 MW; 2DD8B8715B17F74E CRC64;

Query Match 7.5%; Score 76; DB 2; Length 1055;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-67;  
 Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 (1) \_\_\_\_\_  
 QY 805 LVGEKGLAQAATWVGQVYLLKTPPLPEYIYGLNLYFVFKSGKLIYALDMSDTIGSEQKDA 864  
 DB 806 LVGEKGLAQAATWVGQVYLLKTPPLPEYIYGLNLYFVFKSGKLIYALDMSDTIGSEQKDA 865  
 QY 865 YGNPILNVDEDEGHALAVATLADYEG 880  
 DB 866 YGNPILNVDEDEGHALAVATLADYEG 881  
 RESULT 7  
 Q9ZRT5 PRELIMINARY; PRT; 245 AA.  
 AC Q9ZRT5;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Glutathione transferase AtGST 10 (EC 2.5.1.18).  
 GN Name=gat10; Synonyms=MERF-28;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN (1) \_\_\_\_\_  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=whole plant;  
 RA Dixon D.P., Cole D.J., Edwards R.;  
 RT "Identification and cloning of AtGST 10 (Accession Nos. AJ131580 and  
 RT AJ132398), members of a novel type of plant glutathione  
 RT transferases.";  
 RL Plant Physiol. 119:1568-1568 (1999).  
 RN (2) \_\_\_\_\_  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98290546; PubMed=9628582;  
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.  
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen  
 RT physically assigned P1 and TAC clones.";  
 RL DNA Res. 5:41-54 (1998).  
 RN (3) \_\_\_\_\_  
 RP SEQUENCE FROM N.A.  
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN (4) \_\_\_\_\_  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Belongs to the GST superfamily.  
 DR EMBL; AJ131580; CAA10457.1; -;  
 DR EMBL; AB010072; BAB09723.1; -;  
 DR EMBL; AY054659; AAK96850.1; -;  
 DR EMBL; AY072466; AAL66881.1; -;  
 DR PIR; T51594; T51594.  
 DR HSP; P30712; 1LJR.  
 DR GO; GO:0004364; F:glutathione transferase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR InterPro; IPR004046; GST\_Cterm.  
 DR InterPro; IPR010987; GST\_C\_like.

DR InterPro: IPR004045; GST\_Nterm.  
DR Pfam: PF00043; GST\_C; 1.  
DR Pfam: PF02798; GST\_N; 1.  
DR Transferrase.  
SQ SEQUENCE 245 AA; 27653 MW; 6F16PEI1399E793B7 CRC64;

Query Match 0.9%; Score 9; DB 2; Length 245;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 ABAEAEQQL 937  
DB 133 ABAEAEQQL 141  
|||||

RESULT 8  
Q9NW78 PRELIMINARY; PRT; 279 AA.  
AC Q9NW78  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein FL10244.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
[1]  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Whole embryo;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,  
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
RA Ninomiya K., Tahibashi T., Yamashita H., Murakawa K., Fujimori K.,  
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,  
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
RA Moriya S., Momiyauna H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Ingaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigetani K., Senba T.,  
RA Tsumumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Negase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
RA Nakai K., Tada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs";  
RN Nat. Genet. 36:40-45(2004).  
DR EMBL; AK001106; BAA91506.1; -;  
DR GO; GO:0005085; F:guanylyl-nucleotide exchange factor activity; IEA.  
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.  
DR InterPro: IPR001895; RasGEF CDC25.  
DR InterPro: IPR008937; Ras GEF.  
DR Pfam: PF00617; RasGEF; 1.  
DR SMART: SM00147; RasGEF; 1.  
DR PROSITE: PS50009; RasGEF CAT; 1.  
SQ SEQUENCE 279 AA; 31856 MW; 92195039C78F56EA CRC64;

Query Match 0.9%; Score 9; DB 2; Length 279;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSSES 952  
DB 88 SEKSSSSSES 96  
|||||

RESULT 9  
Q86A18 PRELIMINARY; PRT; 455 AA.  
AC Q86A18  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Similar to Homo sapiens (Human). Dentin sialoposphoprotein.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OC NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX4;  
RX MEDLINE=22092622; PubMed=12097910; DOI=10.1038/nature00847;  
RA Glockner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";  
RL Nature 418:79-85(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX4;  
RX Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; ACL117176; AAO52137.1; -; DB000463D1FEA2A4 CRC64;  
SQ SEQUENCE 455 AA; 51270 MW; DBB00463D1FEA2A4 CRC64;

Query Match 0.9%; Score 9; DB 2; Length 455;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSSES 952  
DB 88 SEKSSSSSES 96  
|||||

RESULT 10  
Q8C134 PRELIMINARY; PRT; 523 AA.  
AC Q8C134  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched  
DE library, clone:4732485D01 product:RAL-A EXCHANGE FACTOR RALGPS2  
DE homolog.  
GN Name=Ralgs2;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OC NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Skin;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Meth. Enzymol. 303:19-44(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Skin;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA RIKEN FANTOM Consortium;  
RT "Functional annotation of a full-length mouse cDNA collection.";

```
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
  60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
  prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki K.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
  sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kaikawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Nunazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AK029049; BAC26264.1; -.
DR MGD; MGI:1925505; Ralgs2.
DR GO; GO:0005085; F:guanylyl-nucleotide exchange factor activity; IEA.
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR InterPro; IPR001895; RasGRF CDC25.
DR InterPro; IPR008937; Ras_GEF.
DR Pfam; PF00617; RasGEF; 1.
DR SMART; SM00147; RasGEF; 1.
DR PROSITE; PS50009; RasGEF CAT; 1.
SQ SEQUENCE 523 AA; 57863 MW; D89582752DBA9F4D CRC64;

Query Match 0.9%; Score 9; DB 2; Length 523;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSES 952
DB 18 SEKSSSES 26
|||||

RESULT 11
Q80YA6 PRELIMINARY; PRT; 564 AA.
ID Q80YA6;
AC Q80YA6;
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE library, clone:9130014M22 product:RAL-A EXCHANGE FACTOR RALGFS2
DE homolog.
GN Name=4921528G01Rik;

01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Ralgs2 protein.
GN Name=Ralgs2;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Shevchenko Y., Bouffard G.G.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
  and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBSJ databases.
CC -!- SIMILARITY: Contains 1 PH domain.
DR EMBL; BC043132; AAH43132.1; -.
DR MGD; MGI:1925505; Ralgs2.
DR GO; GO:0005085; F:guanylyl-nucleotide exchange factor activity; IEA.
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH related.
DR InterPro; IPR001895; RasGRF CDC25.
DR InterPro; IPR008937; Ras_GEF.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00617; RasGEF; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00147; RasGEF; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50009; RasGEF CAT; 1.
SQ SEQUENCE 564 AA; 62524 MW; A2988CD76323F0EE CRC64;

Query Match 0.9%; Score 9; DB 2; Length 564;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSES 952
DB 18 SEKSSSES 26
|||||

RESULT 12
Q9D2Y7 PRELIMINARY; PRT; 568 AA.
ID Q9D2Y7;
AC Q9D2Y7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male cecum cDNA, RIKEN full-length enriched
DE library, clone:9130014M22 product:RAL-A EXCHANGE FACTOR RALGFS2
DE homolog.
GN Name=4921528G01Rik;
```





```

Db          18 SEKSSSES 26
RESULT 14
Q9FHE1      PRELIMINARY;      PRT;      590 AA.
AC Q9FHE1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glutathione transferase-like.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eusoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20181125; PubMed=10718197;
RX Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:31-63(2000).
CC -1- SIMILARITY: Contains 1 Myb-like domain.
DR EMBL; AB019325; BAB11098.1; -.
DR HSP; Q9ZVQ3; 1E6B.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR010987; GST_C-like.
DR InterPro; IPR004045; GST_Nterm.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR PROSITE; PS50090; MYB_3; 1.
KW Transferase.
SQ SEQUENCE 590 AA; 67554 MW; E85F6525F6C7AF5 CRC64;

Query Match          0.9%; Score 9; DB 2; Length 590;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAAEAFQLL 937
Db 132 AAAAEAFQLL 140

RESULT 15
Q7TP26      PRELIMINARY;      PRT;      590 AA.
AC Q7TP26;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Ralgs2 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Egg;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

```

```

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL; BC052663; AAHS2663.1; -.
DR GO; GO:0005085; P:guanyl-nucleotide exchange factor activity; IEA.
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR InterPro; IPR011849; PH.
DR InterPro; IPR011036; PH-related.
DR InterPro; IPR001895; RasGRF_CDC25.
DR InterPro; IPR008937; Ras_GRF.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00617; RasGEF; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00147; RasGEF; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50009; RasGEF_CAT; 1.
SQ SEQUENCE 590 AA; 65458 MW; CAC5CD67FF04DE0F CRC64;

Query Match          0.9%; Score 9; DB 2; Length 590;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSES 952
Db 18 SEKSSSES 26

RESULT 16
Q8BZU2      PRELIMINARY;      PRT;      590 AA.
AC Q8BZU2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus adult male colon cDNA, RIKEN full-length enriched
DE library, clone:9030616022 product:RAL-A EXCHANGE FACTOR RALGPS2
DE homolog.
GN Name=Ralgs2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Colon;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;

```

RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Colon;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Colon;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Colon;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kusunagi T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Colon;  
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
RA Saito K., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC -/- SIMILARITY: Contains 1 PH domain.  
DR EMBL; AK033549; BAC28351.1; -.  
DR MGD; MGI:1925505; RalGPS2.  
DR GO; GO:0005085; P:guanyl-nucleotide exchange factor activity; IEA.  
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR011036; PH-related.  
DR InterPro; IPR001895; RasGRP CDC25.  
DR Pfam; PF00169; PH; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00147; RasGEF; 1.  
DR PROSITE; PS50003; PH DOMAIN; 1.  
DR PROSITE; PS50009; RasGEF CAT; 1.  
SQ SEQUENCE 590 AA; 65544 MW; 3F337D0F359A4DB3 CRC64;

Query Match 0.9%; Score 9; DB 2; Length 590;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 944 SEKSSSES 952  
DB 18 SEKSSSES 26

RESULT 17  
Q9ERD6 PRELIMINARY; PRT; 590 AA.  
AC Q9ERD6;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Ral-A exchange factor RalGPS2.  
GN Name=RalGPS2;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c; TISSUE=Testis;  
RA Martegani E., Cariani M., Bossi D.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
CC -/- SIMILARITY: Contains 1 PH domain.  
DR EMBL; AF312924; AAG34162.1; -.  
DR MGD; MGI:1925505; RalGPS2.  
DR GO; GO:0005085; P:guanyl-nucleotide exchange factor activity; IEA.  
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR011036; PH related.  
DR InterPro; IPR001895; RasGRP CDC25.  
DR InterPro; IPR008937; Ras\_GEF.  
DR Pfam; PF00169; PH; 1.  
DR Pfam; PF00617; RasGEF; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00147; RasGEF; 1.  
DR PROSITE; PS50003; PH DOMAIN; 1.  
DR PROSITE; PS50009; RasGEF CAT; 1.  
SQ SEQUENCE 590 AA; 65548 MW; PF52DE979A745EF2 CRC64;  
Query Match 0.9%; Score 9; DB 2; Length 590;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 944 SEKSSSES 952  
DB 18 SEKSSSES 26  
RESULT 18  
LGR5 HUMAN  
ID LGR5 HUMAN STANDARD; PRT; 907 AA.  
AC Q75473; Q9UP75;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Leucine-rich repeat-containing G protein-coupled receptor 5 precursor  
DE (Orphan G protein-coupled receptor HG38) (G protein-coupled receptor  
DE 49).  
GN Name=GPR49; Synonyms=GPR67, LGR5;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98308104; PubMed=9642114; DOI=10.1006/bbrc.1998.8774;  
RA McDonald T., Wang R., Bailey W., Xie G., Chen F., Caskey C.T., Liu Q.;  
RT "Identification and cloning of an orphan G protein-coupled receptor of  
the glycoprotein hormone receptor subfamily.";  
RL Biochem. Biophys. Res. Commun. 247:266-270(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=99065210; PubMed=9849958; DOI=10.1210/me.12.12.1830;  
RA Hsu S.Y., Liang S.-G., Haueh A.J.W.;  
RT "Characterization of two LGR genes homologous to gonadotropin and



```

RN (4)
RP SEQUENCE FROM N.A.
RA Wilson R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AC004884; AAS00352.1; -.
KW Hypothetical protein.
FT NON_TER 12 12
SQ SEQUENCE 12 AA; 1271 MW; 98816D259CB33B05 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSE 951
DB 5 SEKSSSSE 12
|||||
|||||

RESULT 20
Q982N8 PRELIMINARY; PRT; 67 AA.
AC Q982N8;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Mr8560 protein.
GN OrderedLocNames=mr8560;
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideawara K., Iehikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP03014; BAB54418.1; -.
KW Complete proteome.
SQ SEQUENCE 67 AA; 7250 MW; 27F1924C9BFC93E5 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 67;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 757 KALVTKKA 764
DB 52 KALVTKKA 59
|||||
|||||

RESULT 21
Q7UVJ9 PRELIMINARY; PRT; 98 AA.
AC Q7UVJ9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Probable ethanolamine utilization protein EutM.
GN Name=eutM; OrderedLocNames=RB2586;
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1;
RL MEDLINE=22735913; PubMed=12835416; DOI=10.1073/pnas.1431443100;

RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; BX294137; CAD72725.1; -.
DR InterPro; IPR000249; Bact_microcomp.
DR Pfam; PF00936; Bact_microcomp; 1.
DR ProDom; PD003442; Bact_microcomp; 1.
KW Complete proteome.
SQ SEQUENCE 98 AA; 9953 MW; 60966543693CC8C2 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 687 DLEKILKL 694
DB 85 DLEKILKL 92
|||||
|||||

RESULT 22
RL3E_SULTO
ID _RL3E_SULTO STANDARD; PRT; 106 AA.
AC P58376;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE 50S ribosomal protein L30e.
GN Name=rlp30e; OrderedLocNames=ST0274;
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OX Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JCM 10545 / 7;
MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7."
RL DNA Res. 8:123-140(2001).
CC -1- SIMILARITY: Belongs to the ribosomal protein L30e family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000982; BAB65241.1; -.
DR HSSP; P29160; 1H7M.
DR HAMAP; MF_00481; -.
DR InterPro; IPR000231; Ribosomal_L30e.
DR InterPro; IPR004038; Ribosomal_L7A.
DR Pfam; PF01248; Ribosomal_L7Ae; 1.
DR ProDom; PD004495; Ribosomal_L7A; 1.
DR PROSITE; PS00709; RIBOSOMAL_L30E_1; 1.
DR PROSITE; PS00993; RIBOSOMAL_L30E_2; 1.
KW Complete proteome; Ribosomal protein.
SQ SEQUENCE 106 AA; 11616 MW; DB9F4D86E5CF57CEF CRC64;

Query Match 0.8%; Score 8; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

factors such as greA or greB allows the resumption of elongation from the new 3'-terminus (By similarity)).

-1- SIMILARITY: Belongs to the greA/greB family.

EMBL; AE016872; AAO58047.1; -.

DR HSSP; P21346; 1GRJ.

DR TIGR; PSPT04601; -.

DR GO; GO:0003677; F:DNA binding; IEA.

DR GO; GO:0003711; F:transcriptional elongation regulator activity; IEA.

DR GO; GO:0003746; F:translation elongation factor activity; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro; IPR001437; GreA\_Greb.

DR Pfam; PF01272; GreA\_Greb; 1.

DR ProDom; PD004918; GreA\_Greb; 1.

KW Complete proteome; DNA-binding; Elongation factor; Transcription; Transcription regulation.

SQ SEQUENCE 165 AA; 18080 MW; 7A47CC37PD178B58 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 165;

Best Local Similarity 100.0%; Pred.No. 76;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 928 NAAAEAEQ 935

DB 10 NAAAEAEQ 17

|||||

RESULT 25

Q9SB23 PRELIMINARY; PRT; 174 AA.

AC Q9SB23

DT 01-MAY-2000 (TREMELrel. 13, Created)

DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)

DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)

DE Beta-amylase (EC 3.2.1.2) (Fragment).

GN Name=beta-Amy.

OS Hordeum vulgare (Barley).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;

OC Triticeae; Hordeum.

OX NCBI\_TaxID=4513;

RN [1]\_TaxID=4513;

RP SEQUENCE FROM N.A.

RC STRAIN=Haruna NiJo;

RA Okada Y., Kihara M., Kuroda H., Yoshigi N., Ito K.;

RT "Cloning and Sequencing of the Promoter Region of the Seed Specific beta-Amylase Gene from Barley";

RL J. Plant Physiol. 156:762-767(2000).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in polysaccharides so as to remove successive maltose units from the non-reducing ends of the chains.

CC -1- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.

DR EMBL; D63574; BAA09793.1; -.

DR HSSP; P16098; 1BY1.

DR GO; GO:0016161; F:beta-amylase activity; IEA.

DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR GO; GO:0000272; P:polysaccharide catabolism; IEA.

DR InterPro; IPR001554; Glyco\_hydro\_14.

DR Pfam; PF01373; Glyco\_hydro\_14; 1.

DR PRINTS; PR00750; BETAAMYLASE.

DR PROSITE; PS00506; BETA\_AMYLASE 1; 1.

KW Carbohydrate metabolism; Glycosidase; Hydrolase;

KW Polysaccharide degradation.

FT NON\_TER 174 174

SQ SEQUENCE 174 AA; 19593 MW; C8APAG1737153279 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 174;

Best Local Similarity 100.0%; Pred.No. 80;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEKGKP 811

DB 55 GLVEKGKP 62

|||||

501 LEBIKEGA 508

|||||

98 LEBIKEGA 105

RESULT 23

Q9NS28 PRELIMINARY; PRT; 116 AA.

AC Q9NS28

DT 01-OCT-2000 (TREMELrel. 15, Created)

DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)

DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)

DE Hypothetical protein DKFZp34H0512 (Fragment).

GN Name=DKFZp34H0512;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RA Duesterhoeft A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL137620; CAB70846.1; -.

DR PIR; T46320; T46320.

KW Hypothetical protein.

FT NON\_TER 1 1

SQ SEQUENCE 116 AA; 13372 MW; CED73BD670665BE2 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 116;

Best Local Similarity 100.0%; Pred.No. 55;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 930 AABAEQLL 937

DB 83 AABAEQLL 90

|||||

RESULT 24

Q87WF0 PRELIMINARY; PRT; 165 AA.

AC Q87WF0

DT 01-JUN-2003 (TREMELrel. 24, Created)

DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Transcription elongation factor GreB

GN Name=greB-2; OrderedLocustNames=PSPT04601;

OS Pseudomonas syringae (pv. tomato).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=323;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DC3000;

RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;

RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T., Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H., Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q., Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J., Uterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M., Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K., Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X., Bender C.L., White O., Fraser C.M., Collier A.;

RT "The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000.";

RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).

CC -1- FUNCTION: Necessary for efficient RNA polymerase transcription elongation past template-encoded arresting sites. The arresting sites in DNA have the property of trapping a certain fraction of elongating RNA polymerases that pass through, resulting in locked ternary complexes. Cleavage of the nascent transcript by cleavage



```

RESULT 26
Q9FR02
ID Q9FR02 PRELIMINARY; PRT; 237 AA.
AC Q9FR02
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE AV-9/Cf-9 rapidly elicited protein 1.
GN Name=ACRE1;
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20312759; PubMed=10852940;
RA Durrant W.B., Rowland O., Piedras P., Hammond-Kosack K.E., Jones J.D.;
RT "cDNA-AFLP reveals a striking overlap in race-specific resistance and
RT wound response gene expression profiles.";
RL Plant Cell 12:963-977(2000).
DR EMBL; AF211527; AAG43545.1; -.
DR HSP; O80337; 2GCC.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003709; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001471; TF_ERF.
DR PRINTS; PR00367; ETRHSPLEWNT.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
SQ SEQUENCE 237 AA; 26513 MW; 823E51544327BB87 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 948 SSSSANS 955
DB 42 SSSSANS 49

RESULT 27
Q8TW63
ID Q8TW63 PRELIMINARY; PRT; 240 AA.
AC Q8TW63
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Uncharacterized protein specific for M.kandleri, MK-27 family.
GN OrderedLocusNames=Mk1173;
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus;
OX NCBI_TaxID=23320;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AV19 / DSM 6324 / JCM 9639;
RY MEDLINE=21927647; PubMed=11930014; DOI=10.1073/pnas.032671499;
RA Slesarev A.I., Mezhevaeva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Nalele D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozyavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL; AE010408; AM02386.1; -.
KW Complete proteome.
SQ SEQUENCE 240 AA; 25400 MW; 59813D9792655BEC CRC64;

Query Match 0.8%; Score 8; DB 2; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 948 SSSSANS 955
DB 42 SSSSANS 49

RESULT 28
Q7T3A2
ID Q7T3A2 PRELIMINARY; PRT; 241 AA.
AC Q7T3A2
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Zgc:64012.
GN ORFNames=zgc:64012;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053197; AAHS3197.1; -.
DR ZFIN; ZDB-GENE-040426-2639; zgc:64012.
DR GO; GO:0005525; P:GTP binding; IEA.
SQ SEQUENCE 241 AA; 26617 MW; C625FE64088D3D98 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 EKIESEAP 356
DB 198 EKIESEAP 205

RESULT 29
Q7NST9
ID Q7NST9 PRELIMINARY; PRT; 242 AA.
AC Q7NST9
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Probable ABC transporter, ATP-binding protein.
GN OrderedLocusNames=CV3331;

```

OS Chromobacterium violaceum.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Chromobacterium.  
 OX NCBI\_TaxID=536;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12472 / DSM 30191;  
 RX MEDLINE=2282880; PubMed=14500782; DOI=10.1073/pnas.1032124100;  
 RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,  
 RA Antonio R.V., Almeida F.C. de Almeida L.G.P., de Almeida R.,  
 RA Alves-Gomes J.A., Andrade E.M., Aralide J., de Araujo M.F.F.,  
 RA Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,  
 RA Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,  
 RA Bordignon J., Brigido M.M., Brito C.A., Brocchi M., Buriti H.A.,  
 RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,  
 RA Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,  
 RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,  
 RA Faninatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,  
 RA Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,  
 RA Grattapaglia D., Griesard E.C., Hanna E.S., Jardim S.N., Laurino J.,  
 RA Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,  
 RA Madeira H.M.F., Manfio G.P., Maranhao A.Q., Martins W.S.,  
 RA di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,  
 RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,  
 RA Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,  
 RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,  
 RA Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,  
 RA Santos E.B.P., Santos F.R., Schneider M.P.C., Seunanez H.N.,  
 RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,  
 RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,  
 RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,  
 RA Vettore A., Wassen R., Zaha A., Simpson A.J.G.;  
 RT "The complete genome sequence of Chromobacterium violaceum reveals  
 RT remarkable and exploitable bacterial adaptability";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665 (2003).  
 RC -!- SIMILARITY: Belongs to the ABC transporter family.  
 DR EMBL; AB016921; AAQ60995.1; -;  
 DR HSP; 058663; 1G9X.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.  
 DR GO; GO:0042829; P:defense response to pathogen; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR InterPro: IPR000767; Disease\_resist.  
 DR InterPro: IPR003016; Lipoyl\_BS.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR PRINTS; PR00364; DISEASERSIST.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00189; LIPOYL; UNKNOWN\_1.  
 KW ATP-binding; Complete proteome.  
 SQ SEQUENCE 242 AA; 26822 MW; 912D717124132EA8 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 242;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 LEKELDLL 802  
 Db 115 LEKELDLL 122  
 |||||

RESULT 30  
 P93664 PRELIMINARY; PRT; 253 AA.  
 AC P93664;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE LI818r-1 protein precursor.

GN Name=LI818r-1;  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 OX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CC620;  
 RA Guertin M.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X95326; CAA64632.1; -;  
 DR PIR; T08175; T08175.  
 DR GO; GO:0009765; P:photosynthesis light harvesting; IEA.  
 DR InterPro: IPR001344; Chloro Abbind.  
 DR Pfam; PF00504; Chloroa\_b-bind; 1.  
 KW Signal.  
 FT SIGNAL 1 32 Potential.  
 FT CHAIN 33 253 Potential.  
 SQ SEQUENCE 253 AA; 27556 MW; 985274D58C5BBF1D CRC64;

Query Match 0.8%; Score 8; DB 2; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 LPDNLKAI 283  
 Db 246 LPDNLKAI 253  
 |||||

RESULT 31  
 Q9AB78 PRELIMINARY; PRT; 255 AA.  
 ID Q9AB78  
 AC Q9AB78;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocustNames=BT0232;  
 OS Bacteroides thetaiotaomicron.  
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;  
 OC Bacteroidaceae; Bacteroides.  
 OX NCBI\_TaxID=818;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VPI-5482 / ATCC 29148;  
 RX MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;  
 RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,  
 RA Chiang H.C., Hooper L.V., Gordon J.I.;  
 RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis";  
 RL Science 299:2074-2076 (2003).  
 DR EMBL; AE016926; AA075339.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 255 AA; 29434 MW; 4894A4FC5168A8E2 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 255;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 696 EGLDYSTL 703  
 Db 215 EGLDYSTL 222  
 |||||

RESULT 32  
 Q8UYE5 PRELIMINARY; PRT; 255 AA.  
 ID Q8UYE5  
 AC Q8UYE5;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Putative chemokine binding protein.  
 OS Camelpox virus (strain M-96).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

```

OC Orthopoxvirus.
OX NCBI_TaxID=203173;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22031204; PubMed=12033760; DOI=10.1006/viro.2001.1343;
RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Sandybaev N.T.,
RT Kerembekova U.Z., Zaitsev V.L., Kutish G.F., Rock D.L.;
RT "The genome of camelpox virus.";
RL Virology 295:1-9(2002).
DR EMBL; AF438165; AAL73919.1; -.
DR EMBL; AF438165; AAL73919.1; -.
DR HSFP; O73568; ICQ3.
DR PFam; PF02250; Orthopox_35kd; 1.
DR PIRSF; PIRSF003696; VAC_C23L; 1.
SQ SEQUENCE 255 AA; 27659 MW; 84B8DF0C01CE31FE CRC64;

Query Match 0.8%; Score 8; DB 2; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 VTENTESE 106
Db 61 VTENTESE 68

RESULT 33
Q8YLZ5 ID Q6YLZ5 PRELIMINARY; PRT; 255 AA.
AC Q6YLZ5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Secreted chemokine binding protein.
OS Camelpox virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=28873;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CP-1, CP-1231 Haut Dubai, CP-1260/95 Haut Dubai, CP-5, and
RC Saudi;
RA Mikheev M.V., Feshchenko M.V., Shchelkunov S.N.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY102980; AAM76329.1; -.
DR EMBL; AY102981; AAM76330.1; -.
DR EMBL; AY102982; AAM76331.1; -.
DR EMBL; AY102984; AAM76333.1; -.
DR EMBL; AY102983; AAM76332.1; -.
DR EMBL; AY102983; AAM76332.1; -.
DR InterPro; IPR003184; Orthopox_35kd.
DR PFam; PF02250; Orthopox_35kd; 1.
DR PIRSF; PIRSF003696; VAC_C23L; 1.
SQ SEQUENCE 255 AA; 27659 MW; 84B8DF0C01CE31FE CRC64;

Query Match 0.8%; Score 8; DB 2; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 VTENTESE 106
Db 61 VTENTESE 68

RESULT 34
Q776C0 ID Q776C0 PRELIMINARY; PRT; 255 AA.
AC Q776C0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE CMF1L (CMP206R).
GN Name=CMF1L; Synonyms=CMF206R;
OS Camelpox virus (strain CMS).

```

```

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=203172;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CMS;
RX PubMed=11907336;
RA Gubser C., Smith G.L.;
RT "The sequence of camelpox virus shows it is most closely related to
RT variola virus, the cause of smallpox.";
RL J. Gen. Virol. 83:855-872(2002).
DR EMBL; AY009089; AAG37455.1; -.
DR EMBL; AY009089; AAG37455.1; -.
DR InterPro; IPR003184; Orthopox_35kd.
DR InterPro; IPR009173; VAC_C23L.
DR PFam; PF02250; Orthopox_35kd; 1.
DR PIRSF; PIRSF003696; VAC_C23L; 1.
SQ SEQUENCE 255 AA; 27659 MW; 84B8DF0C01CE31FE CRC64;

Query Match 0.8%; Score 8; DB 2; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 VTENTESE 106
Db 61 VTENTESE 68

RESULT 35
Q9H0J8 ID Q9H0J8 PRELIMINARY; PRT; 268 AA.
AC Q9H0J8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein DKFZp434B0516.
GN Name=DKFZp434B0516;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Kohrer K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL136767; CAB66701.1; -.
KW Hypothetical protein.
SQ SEQUENCE 268 AA; 31293 MW; BBE2BA96541635D7 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 793 KRLSKELD 800
Db 83 KRLSKELD 90

RESULT 36
Q96191 ID Q96191 PRELIMINARY; PRT; 284 AA.
AC Q96191;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein PFB0490c.
GN Name=PFB0490c;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=99021743; PubMed=9804551; DOI=10.1126/science.282.5391.1126;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalton S., Mason T., Yu K., Fujii C., Pedersen J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perte A.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shalton S.J., Suh B., Peterson J., Angiuoli S.,
RA Perte A., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE001398; AAC71887.1; -.
DR PIR; E71613; E71613.
KW Hypothetical protein.
SQ SEQUENCE 284 AA; 32541 MW; 41B35EFBC2C82B43 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 284;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 EVNLPESL 258
DB 269 EVNLPESL 276

RESULT 37
Q926G6 PRELIMINARY; PRT; 296 AA.
AC Q926G6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative sugar uptake ABC transporter periplasmic solute-binding
DE protein.
GN ORFNames=SWB21377;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431; DOI=10.1073/pnas.161294698;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoelter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL591985; CAC49355.1; -.
DR PIR; C95961; C95961.
DR HSSP; P02925; 2DRI.
DR Pfam; PF00532; Peripla_BP_1.1.
KW Complete proteome; Plasmid.
SQ SEQUENCE 296 AA; 30625 MW; 43CB86F44829C9F3 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 296;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 751 AIAKAEKA 758
DB 98 AIAKAEKA 105

RESULT 38
Q9ABZ6 PRELIMINARY; PRT; 304 AA.
AC Q9ABZ6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein CC0074.
GN OrderedLocustNames=CC0074;
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647; DOI=10.1073/pnas.061029298;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H.M., Shetty J.,
RA Berry K.J., Utterback T.R., Tran K., Wolf A.M., Vamathevan J.J.,
RA Ermolaeva M.D., White O., Salzberg S.L., Venter J.C., Shapiro L.,
RA Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005681; AAK22061.1; -.
DR PIR; A87258; A87258.
DR TIGR; CC0074; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 304 AA; 31526 MW; EDOA35B28AD493E1 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 881 ALAVATLA 888
DB 131 ALAVATLA 138

RESULT 39
Q88U35 PRELIMINARY; PRT; 316 AA.
AC Q88U35;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein lp.2713.
GN OrderedLocustNames=lp.2713;
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12565656; DOI=10.1073/pnas.0337704100;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL; AL935260; CAD64951.1; -.

```

KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 316 AA; 34797 MW; E9DB4EA1AF03776D CRC64;  
Query Match 0.8%; Score 8; DB 2; Length 316;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 630 LEVASALS 637  
Db 227 LEVASALS 234  
RESULT 40  
YDZA SCHPO STANDARD; PRT; 329 AA.  
AC 013717;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Hypothetical protein C14C4.10c in chromosome I.  
GN ORFName=SPAC14C4.10c;  
OS Schizosaccharomyces pombe (Pission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;  
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Welljens I., Vansreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Frizic C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallardrin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Chuzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT The genome sequence of Schizosaccharomyces pombe.  
RL Nature 415:871-880(2002).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; Z98596; CAB11203.1; -.  
DR PIR; T37695; T37695.  
DR GenBank; SPAC14C4.10c; -.  
DR InterPro; IPR000086; NUDIX\_hydrolase.  
DR Pfam; PF00293; NUDIX; 1.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 123 143 Potential.

FT TRANSMEM 227 247 Potential.  
FT TRANSMEM 303 323 Potential.  
SQ SEQUENCE 329 AA; 38293 MW; 5DB9C067BA1E1743 CRC64;  
Query Match 0.8%; Score 8; DB 1; Length 329;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 798 ELDLLTGL 805  
Db 4 ELDLLTGL 11  
RESULT 41  
O99927  
ID O99927 PRELIMINARY; PRT; 348 AA.  
AC O99927;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE NADH dehydrogenase subunit 2.  
GN Name=ND2;  
OS Cyprinella lutrensis (Red shiner).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Cyprinella.  
OX NCBI\_TaxID=28791;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Broughton R.E., Gold J.R.;  
RT "Phylogenetic relationships in the North American cyprinid genus  
RT Cyprinella (Actinopterygii: Cyprinidae) based on sequences of the  
RL mitochondrial ND2 and ND4L genes."  
RL Copeia 2000:1-10(2000).  
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
CC inner membrane (By similarity).  
CC -1- SIMILARITY: Belongs to the complex I subunit 2 family.  
CC EMBL; AF111211; AAD13235.1; -.  
DR GO; GO:0005739; C:mitochondrion; IEA.  
DR GO; GO:0008137; P:NADH dehydrogenase (ubiquinone) activity; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.  
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.  
DR InterPro; IPR003917; NADH dehydrog. S2 C.  
DR InterPro; IPR01750; Oxidored. q1.  
DR Pfam; PF06444; NADH dehy S2 C7.1.  
DR Pfam; PF00361; Oxidored\_q1.1.  
DR PRINTS; PR01436; NADHGNASE2.  
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.  
SQ SEQUENCE 348 AA; 37586 MW; 377C3A14A3E50625 CRC64;  
Query Match 0.8%; Score 8; DB 2; Length 348;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 799 LDLLTGLV 806  
Db 122 LDLLTGLV 129  
RESULT 42  
O99944  
ID O99944 PRELIMINARY; PRT; 348 AA.  
AC O99944;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE NADH dehydrogenase subunit 2.  
GN Name=ND2;  
OS Erimonax monachus.

OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;  
 OC Cyprinidae; Erimonax.  
 OX NCBI\_TaxID=87719;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Broughton R.E., Gold J.R.;  
 RT "Phylogenetic relationships in the North American cyprinid genus  
 RT *Cyprinella* (Actinopterygii: Cyprinidae) based on sequences of the  
 RT mitochondrial ND2 and ND4L genes";  
 RL Copela 2000:1-10(2000).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC inner membrane (By similarity).  
 CC -1- SIMILARITY: Belongs to the complex I subunit 2 family.  
 DR EMBL, AF111228; RAD13252.1; -.  
 DR GO; GO:0005739; C:mitochondrion; IEA.  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.  
 DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . . ; IEA.  
 DR InterPro; IPR003917; NADhub\_oxred2.  
 DR InterPro; IPR010933; NADH\_dchy\_S2\_C.  
 DR Pfam; PF06444; NADH\_dchy\_S2\_C; 1.  
 DR PRINTS; PR01436; NADHDHGNASE2.  
 DR Mitochondrion; NAD; Oxidoreductase; Ubiquinone.  
 SQ SEQUENCE 348 AA; 37634 MW; 3417A690E3F4B9A8 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 LDLLTGLV 806  
 DB 122 LDLLTGLV 129  
 |||||

RESULT 43  
 Q8M518 PRELIMINARY; PRT; 348 AA.  
 AC Q8M518;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE NADH dehydrogenase subunit 2.  
 GN Name=ND2;  
 OS Pelvicachromis pulcher (Rainbow krib).  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;  
 OC Cichlidae; Pelvicachromis.  
 OX NCBI\_TaxID=28827;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Klett V., Meyer A.;  
 RT "What, if anything, is a Tilapia?—mitochondrial ND2 phylogeny of  
 RT tilapines and the evolution of parental care systems in the African  
 RT cichlid fishes";  
 RL Mol. Biol. Evol. 19:865-883(2002).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC inner membrane (By similarity).  
 CC -1- SIMILARITY: Belongs to the complex I subunit 2 family.  
 DR EMBL, AF317271; AM20784.1; -.  
 DR GO; GO:0005739; C:mitochondrion; IEA.  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.

DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . . ; IEA.  
 DR InterPro; IPR003917; NADhub\_oxred2.  
 DR InterPro; IPR010933; NADH\_dchy\_S2\_C.  
 DR InterPro; IPR001750; Oxidored\_q1.  
 DR Pfam; PF06444; NADH\_dchy\_S2\_C; 1.  
 DR Pfam; PF00361; Oxidored\_q1; 1.  
 DR PRINTS; PR01436; NADHDHGNASE2.  
 DR Mitochondrion; NAD; Oxidoreductase; Ubiquinone.  
 SQ SEQUENCE 348 AA; 37879 MW; 85509DE660A1906 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 LDLLTGLV 806  
 DB 122 LDLLTGLV 129  
 |||||

RESULT 44  
 Q8HLA4 PRELIMINARY; PRT; 348 AA.  
 AC Q8HLA4;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE NADH dehydrogenase subunit 2.  
 GN Name=ND2;  
 OS Indostomus paradoxus (Armoured stickleback).  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Perciformes; Gasterosteiformes;  
 OC Indostomidae; Indostomus.  
 OX NCBI\_TaxID=181450;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22359432; PubMed=12470944; DOI=10.1016/S1055-7903(02)00332-9;  
 RA Miya M., Takeshima H., Endo H., Ishiguro N.B., Inoue J.G., Mukai T.,  
 RA Sato T.P., Yamaguchi M., Kawaguchi A., Mabuchi K., Shirai S.M.,  
 RA Nishida M.;  
 RT "Major patterns of higher teleostean phylogenies: a new perspective  
 RT based on 100 complete mitochondrial DNA sequences.";  
 RL Mol. Phylogenet. Evol. 26:121-138(2003).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC inner membrane (By similarity).  
 CC -1- SIMILARITY: Belongs to the complex I subunit 2 family.  
 DR EMBL, AP004438; BAC23594.1; -.  
 DR GO; GO:0005739; C:mitochondrion; IEA.  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.  
 DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . . ; IEA.  
 DR InterPro; IPR003917; NADhub\_oxred2.  
 DR InterPro; IPR010933; NADH\_dchy\_S2\_C.  
 DR InterPro; IPR001750; Oxidored\_q1.  
 DR Pfam; PF06444; NADH\_dchy\_S2\_C; 1.  
 DR Pfam; PF00361; Oxidored\_q1; 1.  
 DR PRINTS; PR01436; NADHDHGNASE2.  
 DR Mitochondrion; NAD; Oxidoreductase; Ubiquinone.  
 SQ SEQUENCE 348 AA; 37834 MW; 3E4561B5EB4B3631 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 799 LDLLTGLV 806  
 DB 122 LDLLTGLV 129  
 |||||

RESULT 45



RT Iron assimilation and intracellular infection and contains a strain-specific insertion sequence element.";  
RL Infect. Immun. 70:1842-1852(2002).  
DR EMBL; AF386079; AAM00399.1; --  
DR InterPro; IPR005616; CcmH.  
DR InterPro; IPR002016; Peroxidase.  
DR InterPro; IPR001440; TPR.  
DR InterPro; IPR008941; TPR-like.  
DR Pfam; PF03918; CcmH; 1.  
DR ProDom; PD005662; CcmH; 1.  
DR PROSITE; PS00435; PEROXIDASE\_1; UNKNOWN\_1.  
DR PROSITE; PS50005; TPR; 2.  
DR PROSITE; PS50293; TPR REGION; 1.  
SQ SEQUENCE 360 AA; 41845 MW; 54703067B4317186 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 360;  
Best Local Similarity 100.0%; Pred.No.1.5e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 KAIKAEK 757  
DB 342 KAIKAEK 349  
|||||

RESULT 47  
QF7S2 PRELIMINARY; PRT; 360 AA.

ID QF7S2 PRELIMINARY;  
AC QF7S2;  
DT 01-MAR-2001 (TREMELrel. 16, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DE 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Predicted N-acetylmuramoyl-L-alanine amidase.  
OS Gamma-proteobacterium EBAC31A08.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; environmental samples.  
OX NCBI\_TaxID=133804;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20446260; PubMed=10988064; DOI=10.1126/science.289.5486.1902;  
RA Beja O., Aravind L., Koonin E.V., Suzuki M.T., Hadd A., Nguyen L.P.;  
RA Jovanovich S.B., Gates C.M., Feldman R.A., Spudich J.L., Spudich E.N.,  
RA DeLong E.F.;  
RT "Bacterial rhodopsin: evidence for a new type of phototrophy in the sea."  
RT Science 289:1902-1906(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Beja O., Aravind L., Koonin E.V., Suzuki M.T., Hadd A., Nguyen L.P.,  
RA Jovanovich S.B., Gates C.M., Feldman R.A., DeLong E.F.;  
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF279106; AAG10447.1; --  
DR GO; GO:0008745; F:N-acetylmuramoyl-L-alanine amidase activity; IEA.  
DR GO; GO:0015998; P:cell wall catabolism; IEA.  
DR GO; GO:0009253; P:peptidoglycan catabolism; IEA.  
DR InterPro; IPR002508; Amidase\_3\_hydro.  
DR InterPro; IPR002482; LysM.  
DR Pfam; PF01520; Amidase\_3; 1.  
DR Pfam; PF01476; LysM; 1.  
DR SMART; SM00646; Ami\_3; 1.  
DR SMART; SM00257; LysM; 1.  
SQ SEQUENCE 360 AA; 40939 MW; 3313E148EC6F3C24 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 360;  
Best Local Similarity 100.0%; Pred.No.1.5e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 KVKRKNKL 436  
DB 61 KVKRKNKL 68  
|||||

RESULT 48  
Q87S80 PRELIMINARY; PRT; 370 AA.

```

AC Q87580;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein VP0544.
GN OrderedLocusNames=VP0544;
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:k6;
RX MEDLINE=22508454; PubMed=12620739; DOI=10.1016/S0140-6736(03)12659-1;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Nijima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae."
RL Lancet 361:743-749(2003).
DR EMBL; AP005074; BAC58807.1; -.
DR InterPro; IPR005338; UPF0075.
DR Pfam; PF03702; UPF0075; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 370 AA; 40306 MW; DCSE9DF79049C16E CRC64;

Query Match 0.8%; Score 8; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

QY 279 NLKAIGEL 286
DB 59 NLKAIGEL 66

RESULT 49
Q7QF26
ID Q7QF26 PRELIMINARY; PRT; 385 AA.
AC Q7QF26;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE AgCP13368 (fragment).
GN Name=agCG54736; ORFNames=ENSANGG00000009938;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST.
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008846; EAA06711.1; -.
FT NON_TER 1
SQ SEQUENCE 385 AA; 44555 MW; 3FF4E2A3FB3ABFEE CRC64;

Query Match 0.8%; Score 8; DB 2; Length 385;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 LEIKDKLV 524
DB 211 LEIKDKLV 218

RESULT 50
Q63C90
ID Q63C90 PRELIMINARY; PRT; 391 AA.
AC Q63C90;

```

```

DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Alanine racemase (EC 5.1.1.1).
GN Name=dal; ORFNames=B7ZKI884;
OS Bacillus cereus ZK.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=288681;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZK;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RT "Complete genome sequence of Bacillus cereus ZK."
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000001; AAU18372.1; -.
KW Isomerase.
SQ SEQUENCE 391 AA; 43417 MW; 59CD94ACF3851BD5 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 988 ILGYTVA 995
DB 86 ILGYTVA 93

RESULT 51
Q739H8
ID Q739H8 PRELIMINARY; PRT; 391 AA.
AC Q739H8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Alanine racemase (EC 5.1.1.1).
GN Name=alr; OrderedLocusNames=BCE2163;
OS Bacillus cereus (strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=222523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14960714; DOI=10.1093/nar/gkh258;
RA Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,
RA Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,
RA Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
RT adaptations and a large plasmid related to Bacillus anthracis pX01."
RL Nucleic Acids Res. 32:977-988(2004).
CC -! CATALYTIC ACTIVITY: L-alanine = D-alanine.
CC -! COFACTOR: Pyridoxal phosphate (By similarity).
CC -! SIMILARITY: Belongs to the alanine racemase family.
DR EMBL; AE017271; AAS41084.1; -.
DR HSSP; P10724; 1BD0.
DR TIGR; BCE2163; -.
DR GO; GO:0008784; F:alanine racemase activity; IEA.
DR GO; GO:0030170; F:pyridoxal phosphate binding; IEA.
DR GO; GO:0006522; F:alanine metabolism; IEA.
DR InterPro; IPR000821; Ala_racemase.
DR InterPro; IPR011079; Ala_racemase_C.
DR InterPro; IPR001608; Ala_racemase_N.
DR InterPro; IPR009006; Racem_decarbox_C.
DR Pfam; PF00842; Ala_racemase_C; 1.
DR Pfam; PF01168; Ala_racemase_N; 1.
DR PRINTS; PR00992; ALARACEMASE.
DR TIGRFAMs; TIGR00492; alr; 1.
DR PROSITE; PS00395; ALANINE_RACEMASE; UNKNOWN 1.
KW Complete proteome; Isomerase; Pyridoxal phosphate.
SQ SEQUENCE 391 AA; 43385 MW; .345588619898987 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;

```

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 988 ILGYTSVA 995  
 |||||  
 Db 86 ILGYTSVA 93

RESULT 52

Q81RG8 PRELIMINARY; PRT; 391 AA.  
 AC Q81RG8; Q6H2P3; Q6KTN0;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE Alanine racemase.  
 GN Name=dal-2; OrderedLocusNames=BA2079, BAS1932, GBAA2079;  
 OS Bacillus anthracis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1392;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate Porton;  
 RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;  
 RA Read T.D., Peterson S.N., Toukase N.J., Baillie L.W., Paulsen I.T.,  
 RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
 RA Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,  
 RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,  
 RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
 RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
 RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,  
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
 RA Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,  
 RA Salzberg S.L., Thomson B., Friedlander A.M., Koehler T.M.,  
 RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;  
 RT "The genome sequence of Bacillus anthracis Ames and comparison to  
 RT closely related bacteria.";  
 RL Nature 423:81-86(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate 0581;  
 RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,  
 RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,  
 RA Fraser C.M.;  
 RT "Bacillus anthracis comparative genomics.";  
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBSJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sterne;  
 RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
 RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
 RA Richardson P., Rubin E., Tice H.;  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBSJ databases.  
 CC -!- CATALYTIC ACTIVITY: L-alanine = D-alanine.  
 CC -!- COFACTOR: Pyridoxal phosphate (By similarity).  
 CC -!- SIMILARITY: Belongs to the alanine racemase family.  
 DR EMBL; AE017030; AAP25965.1; -;  
 DR EMBL; AE017334; AAT31196.1; -;  
 DR EMBL; AE017225; AAT54246.1; -;  
 DR HSSP; P10724; 1NIU.  
 DR TIGR; BA2079; -;  
 DR TIGR; GBAA2079; -;  
 DR GO; GO:0008784; F:alanine racemase activity; IEA.  
 DR GO; GO:0016853; F:isomerase activity; IEA.  
 DR GO; GO:0030170; F:pyridoxal phosphate binding; IEA.  
 DR GO; GO:0006522; P:alanine metabolism; IEA.  
 DR InterPro; IPR000821; Ala\_racemase.  
 DR InterPro; IPR011079; Ala\_racemase\_C.  
 DR InterPro; IPR001608; Ala\_racemase\_N.  
 DR InterPro; IPR009006; Racem\_decarbox\_C.  
 DR Pfam; PF00842; Ala\_racemase\_C; 1.  
 DR Pfam; PF01168; Ala\_racemase\_N; 1.  
 DR PRINTS; PR00992; ALARACEMASE.  
 DR TIGRFAMs; TIGR00492; alt; 1.

DR PROSITE; PS00395; ALANINE\_RACEMASE; UNKNOWN 1.  
 KW Complete proteome; Isomerase; Pyridoxal phosphate.  
 SQ SEQUENCE 391 AA; 43371 MW; F8AA173912483DF4 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 391;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 988 ILGYTSVA 995  
 |||||  
 Db 86 ILGYTSVA 93

RESULT 53

Q6HJQ1 PRELIMINARY; PRT; 391 AA.  
 AC Q6HJQ1;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Alanine racemase (EC 5.1.1.1).  
 GN Name=dal; OrderedLocusNames=BT9727\_1894;  
 OS Bacillus thuringiensis (subsp. konkukian).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=180856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=97-27;  
 RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
 RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
 RA Richardson P., Rubin E., Tice H.;  
 RT "Complete genome sequence of Bacillus thuringiensis 97-27.";  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBSJ databases.  
 CC -!- CATALYTIC ACTIVITY: L-alanine = D-alanine.  
 CC -!- COFACTOR: Pyridoxal phosphate (By similarity).  
 CC -!- SIMILARITY: Belongs to the alanine racemase family.  
 DR EMBL; AE017355; AAP53484.1; -;  
 DR GO; GO:0008784; F:alanine racemase activity; IEA.  
 DR GO; GO:0016853; F:isomerase activity; IEA.  
 DR GO; GO:0030170; F:pyridoxal phosphate binding; IEA.  
 DR GO; GO:0006522; P:alanine metabolism; IEA.  
 DR InterPro; IPR000821; Ala\_racemase.  
 DR InterPro; IPR011079; Ala\_racemase\_C.  
 DR InterPro; IPR001608; Ala\_racemase\_N.  
 DR InterPro; IPR009006; Racem\_decarbox\_C.  
 DR Pfam; PF00842; Ala\_racemase\_C; 1.  
 DR Pfam; PF01168; Ala\_racemase\_N; 1.  
 DR PRINTS; PR00992; ALARACEMASE.  
 DR TIGRFAMs; TIGR00492; alt; 1.  
 DR PROSITE; PS00395; ALANINE\_RACEMASE; UNKNOWN 1.  
 KW Complete proteome; Isomerase; Pyridoxal phosphate.  
 SQ SEQUENCE 391 AA; 43419 MW; 2209861B4C733C21 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 391;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 988 ILGYTSVA 995  
 |||||  
 Db 86 ILGYTSVA 93

RESULT 54

ACK2\_RHIME STANDARD; PRT; 393 AA.  
 ID ACK2\_RHIME  
 AC Q9X449; Q9XDG2;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Acetate kinase (EC 2.7.2.1) (Acetokinase).  
 GN Name=ackA;  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=104A14;  
 RX MEDLINE=99194732; PubMed=10094701;  
 RA Summers M.L., Denton M.C., McDermott T.R.;  
 RT "Genes coding for phosphotransacetylase and acetate kinase in  
 RT Sinorhizobium meliloti are in an operon that is inducible by phosphate  
 RT stress and controlled by phoB.";  
 RL J. Bacteriol. 181:2217-2224(1999).  
 RN [2]  
 RP SEQUENCE OF 207-314 FROM N.A.  
 RC STRAIN=104A14;  
 RX MEDLINE=99022212; PubMed=9805396;  
 RA Summers M.L., Elkins J.G., Elliot B.A., McDermott T.R.;  
 RT "Expression and regulation of phosphate stress inducible genes in  
 RT Sinorhizobium meliloti.";  
 RL Mol. Plant Microbe Interact. 11:1094-1101(1998).  
 CC -!- CATALYTIC ACTIVITY: ATP + acetate = ADP + acetyl phosphate.  
 CC -!- PATHWAY: Conversion of acetate to acetyl-CoA; first step.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the acetate kinase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF095903; AAD24358.1; -;  
 DR EMBL; AF074452; RAD42996.1; -;  
 DR HSSP; P38502; IG99.  
 DR HAMAP; MF\_00020; -; 1.  
 DR InterPro; IPR000890; Acetate\_kin.  
 DR InterPro; IPR004372; AcK.  
 DR Pfam; PF00871; Acetate\_kinase; 1.  
 DR PRINTS; PR00471; ACETATEKINASE.  
 DR TIGRfam; TIGR00016; acK; 1.  
 DR PROSITE; PS01075; ACETATE\_KINASE\_1; 1.  
 DR PROSITE; PS01076; ACETATE\_KINASE\_2; FALSE\_NEG.  
 KW Kinase; Transferase.  
 FT CONFLICT 209 212 SGAS -> AEFR (in Ref. 2).  
 FT CONFLICT 227 238 LHRPSTGCRWDT -> GFTALDGLPMGTR (in Ref.  
 FT SEQUENCE 393 AA; 42135 MW; 48FD185524CD6D4C CRC64;  
 Query Match 0.8%; Score 8; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 169 AADGTOLI 176  
 Db |||||  
 47 AADGTOLI 54  
 RESULT 55  
 Q6BJ95 PRELIMINARY; PRT; 416 AA.  
 AC Q6BJ95;  
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Debaryomyces hansenii chromosome G of strain CBS767 of Debaryomyces  
 DE hansenii.  
 GN ORFNames=DEHA0G04763g;  
 OS Debaryomyces hansenii CBS767.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.  
 OX NCBI\_TaxID=284592;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CBST767;  
 RG Genolevures;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E., V.,  
 RA Goffard N., Frangeul L., Aigle M., Anhouard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,  
 RA Boirame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantraye F., Hennequin C., Jaumiaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
 RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,  
 RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., With B.,  
 RA Zeniun-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissbach J.,  
 RA Wincker P., Souciet J.L.;  
 RT "Genome evolution in yeasts.";  
 RL Nature 430:35-44(2004).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBST767;  
 RA Genoscope;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBSJ databases.  
 CC -!- SIMILARITY: Belongs to the AAA ATPase family.  
 DR EMBL; CR382139; CAG90179.1; -;  
 DR GO; GO:0005737; C:cytoplasm; IEA.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0030163; P:protein catabolism; IEA.  
 DR InterPro; IPR005917; 26S\_p45.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003599; AAA\_ATPase\_cent.  
 DR InterPro; IPR003960; AAA\_sub.  
 DR Pfam; PF00004; AAA; 1.  
 DR TIGRfam; TIGR01242; 26Sp45; 1.  
 DR PROSITE; PS00674; AAA; 1.  
 KW ATP-binding.  
 SQ SEQUENCE 416 AA; 46706 MW; 9C4C95550B5F33DE CRC64;  
 Query Match 0.8%; Score 8; DB 2; Length 416;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 795 LEKELDLL 802  
 Db |||||  
 42 LEKELDLL 49  
 RESULT 56  
 AROA\_STRMU STANDARD; PRT; 427 AA.  
 AC Q8DUV8;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-  
 DE enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).  
 GN Name=aroA; OrderedLocusNames=SMU.784;  
 OS Streptococcus mutans;  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=UA159 / ATCC 700610 / Serotype C;  
 RX MEDLINE=2295063; PubMed=12397186; DOI=10.1073/pnas.172501299;  
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,  
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,

```

RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RL pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE014920; AAN58503.1; -.
DR HAMAP; MF_00210; -.
DR InterPro; IPR006264; AroA.
DR InterPro; IPR001986; EPSP synth.
DR Pfam; PF00275; EPSP synthase; 1.
DR ProDom; PD001867; EPSP_synth; 1.
DR TIGRFAMs; TIGR01356; aroA; 1.
DR PROSITE; PS00104; EPSP SYNTHASE 1; 1.
DR PROSITE; PS00885; EPSP SYNTHASE 2; 1.
KW Aromatic amino acid biosynthesis; Complete proteome; Transferase.
SQ SEQUENCE 427 AA; 46088 MW; BF15D9C64DB40AFC CRC64;

Query Match 0.8%; Score 8; DB 1; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 LDVQKEI 214
Db |||||
207 LDVQKEI 214

RESULT 57
Q8CKK5 PRELIMINARY; PRT; 427 AA.
AC Q8CKK5;
DT 01-MAR-2003 [TrEMBLrel. 23, Created]
DT 01-MAR-2003 [TrEMBLrel. 23, Last sequence update]
DT 01-OCT-2003 [TrEMBLrel. 25, Last annotation update]
DE Mus musculus adult male olfactory brain cDNA, RIKEN full-length
DE enriched library, clone:6430703L12 product:hypothetical protein, full
DE insert sequence.
GN Name=Rtnn;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=21085650; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;

```

```

RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RL prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RL sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai C., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tomaru A., Takahashi F., Takaku-Akaira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK032605; BAC27946.1; -.
DR MGI; MGI:2179288; Rtnn.
DR GO; GO:0007368; P:determination of left/right symmetry; IMP.
KW Hypothetical protein.
SQ SEQUENCE 427 AA; 47304 MW; DCF5DB1C96026C9D CRC64;

Query Match 0.8%; Score 8; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 824 LKTPPLP 831
Db |||||
168 LKTPPLP 175

RESULT 58
Q97KA1 PRELIMINARY; PRT; 429 AA.
AC Q97KA1;
DT 01-OCT-2001 [TrEMBLrel. 18, Created]
DT 01-OCT-2001 [TrEMBLrel. 18, Last sequence update]
DT 01-JUN-2003 [TrEMBLrel. 24, Last annotation update]
DE Predicted membrane protein.
GN OrderedLocustName=CAC1018;
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=2139325; PubMed=11466286;  
RX DOI=10.1128/JB.183.16.4823-4838.2001;  
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
RA Tatunov R.L., Sabathe F., Doucette-Stamm L.A., Soucaille P.,  
RA Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing  
RT bacterium Clostridium acetobutylicum";  
RL J. Bacteriol. 183:4823-4838(2001).  
DR EMBL; AE007617; AAK78994.1; -;  
DR PIR; G97025; G97025.  
DR InterPro; IPR011041; Quino\_gluc\_DH.  
KW Complete proteome.  
SQ SEQUENCE 429 AA; 48545 MW; 114D12A7262514AA CRC64;  
  
Query Match 0.8%; Score 8; DB 2; Length 429;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 841 FDKSGKLI 848  
|||||||  
Db 239 FDKSGKLI 246  
  
RESULT 59  
Q9NKK1 PRELIMINARY; PRT; 441 AA.  
AC Q9NKK1  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Peroxisomal multifunctional enzyme MFE homolog.  
GN Name=mfeA;  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KAX4;  
RX MEDLINE=22680914; PubMed=12796309; DOI=10.1128/EC.2.3.638-645.2003;  
RA Matsuoka S., Saito T., Kuwayama H., Morita N., Ochiai H., Maeda M.;  
RT "MFE1, a member of the peroxisomal hydroxyacyl coenzyme A  
RT dehydrogenase family, affects fatty acid metabolism necessary for  
RT morphogenesis in Dictyostelium spp.";  
RL Eukaryotic Cell 2:638-645(2003).  
CC -!- SIMILARITY: Belongs to the short-chain dehydrogenases/reductases  
CC (SDR) family.  
DR EMBL; AB042104; BAA94961.1; -;  
DR HSSP; P97852; 1GZ6.  
DR DictyBase; DDB0201628; mfeA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0003498; F:sterol carrier activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR002198; ADH short.  
DR InterPro; IPR002347; Adh\_short\_C2.  
DR InterPro; IPR003033; SCP2.  
DR Pfam; PF00106; adh\_short; 1.  
DR Pfam; PF02036; SCP2; 1.  
DR PRINTS; PR00081; GDHRDH.  
DR PRINTS; PR00080; SDRFAMILY.  
KW Oxidoreductase.  
SQ SEQUENCE 441 AA; 46387 MW; 2CEF3D979D93BCA CRC64;  
  
Query Match 0.8%; Score 8; DB 2; Length 441;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 79 TPEIKDN 86  
|||||||  
Db 259 TPEIKDN 266

RESULT 60  
ID ARLY\_BACHD STANDARD; PRT; 458 AA.  
AC Q9K921;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Argininosuccinate lyase (EC 4.3.2.1) (Argininosuccinase) (ASAL).  
GN Name=argH; OrderedLocusNames=BH3186;  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132; DOI=10.1093/nar/28.21.4317;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
CC -!- CATALYTIC ACTIVITY: N(omega)-L-arginino)succinate = fumarate + L-  
CC arginine.  
CC -!- PATHWAY: Arginine biosynthesis; eighth (last) step.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -!- SIMILARITY: Belongs to the lyase 1 family. Argininosuccinate lyase  
CC subfamily.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AP001518; BAB06905.1; -;  
DR PIR; B84048; B84048.  
DR HSSP; P24058; 1DCN.  
DR HAMAP; MF\_00006; -; 1.  
DR InterPro; IPR009049; argH.  
DR InterPro; IPR003031; D\_crySTALLIN.  
DR InterPro; IPR000362; Fumarate lyase.  
DR InterPro; IPR008948; L-Aspartase-like.  
DR Pfam; PF02026; Lyase\_1; 1.  
DR PRINTS; PR00145; DCRYSTALLIN.  
DR PRINTS; PR00149; FUMRATELYASE.  
DR TIGRFAMs; TIGR00838; argH; 1.  
DR PROSITE; PS00163; FUMARATE LYASES; 1.  
KW Arginine biosynthesis; Complete proteome; Lyase.  
SQ SEQUENCE 458 AA; 51104 MW; CF8F32B39DF4FC40 CRC64;  
  
Query Match 0.8%; Score 8; DB 1; Length 458;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 750 KAIKAEK 757  
|||||||  
Db 447 KAIKAEK 454  
  
RESULT 61  
Q7VVB6 PRELIMINARY; PRT; 463 AA.  
AC Q7VVB6;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Exodeoxyribonuclease large subunit (EC 3.1.11.6).  
GN Name=xseA; OrderedLocusNames=Bp2762;  
OS Bordetella pertussis.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;



```

OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitch E., Rutter S., Sanders M., Saunders R., Squares S., Stevens K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640419; CAE3037.1; -.
DR GO; GO:0009318; C:exodeoxyribonuclease VII complex; IEA.
DR GO; GO:0008855; F:exodeoxyribonuclease VII activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0006308; P:DNA catabolism; IEA.
DR GO; GO:0003753; Exonuc VII L.
DR InterPro; IPR003753; Exonuc VII L.
DR InterPro; IPR008994; Nucleic acid OB.
DR Pfam; PF01336; tRNA_anti. 1.
DR TIGRFAMs; TIGR00237; xseA; 1.
DR Complete proteome; Hydrolase.
KW Complete proteome; Hydrolase.
SQ SEQUENCE 463 AA; 5024 MW; AAFB2CCER494AD79 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 463;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 768 QOLLERSI 775
DB 27 QOLLERSI 34

RESULT 62
ID Q7W7F4 PRELIMINARY; PRT; 463 AA.
AC Q7W7F4
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Exodeoxyribonuclease large subunit (EC 3.1.11.6).
GN Name=xseA; OrderedLocNames=BPP2566;
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitch E., Rutter S., Sanders M., Saunders R., Squares S., Stevens K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640443; CAE32508.1; -.
DR GO; GO:0009318; C:exodeoxyribonuclease VII complex; IEA.
DR GO; GO:0008855; F:exodeoxyribonuclease VII activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0006308; P:DNA catabolism; IEA.
DR InterPro; IPR003753; Exonuc VII L.
DR InterPro; IPR008994; Nucleic acid OB.
DR Pfam; PF01336; tRNA_anti. 1.
DR TIGRFAMs; TIGR00237; xseA; 1.
DR Complete proteome; Hydrolase.
KW Complete proteome; Hydrolase.
SQ SEQUENCE 463 AA; 5026 MW; 4CC6962BFD491BA4 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 463;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 768 QOLLERSI 775
DB 27 QOLLERSI 34

RESULT 62
ID Q7W7F4 PRELIMINARY; PRT; 463 AA.
AC Q7W7F4
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Exodeoxyribonuclease large subunit (EC 3.1.11.6).
GN Name=xseA; OrderedLocNames=BPP2566;
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitch E., Rutter S., Sanders M., Saunders R., Squares S., Stevens K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640430; CAE37860.1; -.

```

```

RESULT 64
Q9YE79 PRELIMINARY; PRT; 478 AA.
AC Q9YE79;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein APE0691.
GN OrderedLocNames=APE0691;
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AF000060; BAA79667.1; -.
DR PIR; C72658; C72658.
DR InterPro; IPR001992; Bact_sec_systII.
DR Pfam; PF00482; GSP11.F; 1.
SQ Complete proteome; Hypothetical protein.
QY SEQUENCE 478 AA; 49725 MW; F05565A05C2A3F58 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 478;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 LKVIGEAS 333
DB 406 LKVIGEAS 413
|||||

RESULT 65
Q9NW94 PRELIMINARY; PRT; 483 AA.
AC Q9NW94;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ10206.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryo;
RX PubMed=14702039; DOI=10.1038/ngl1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hirakawa K., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

```

```

RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsumawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
RL EMBL; AK001068; BAA91489.1; -.
SQ SEQUENCE 483 AA; 54325 MW; 86CABC2ED5C76176 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 945 EKSSSES 952
DB 444 EKSSSES 451
|||||

RESULT 66
Q6LUN0 PRELIMINARY; PRT; 485 AA.
AC Q6LUN0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative sigma-54 dependent transcriptional regulator.
GN Name=CPSR; OrderedLocNames=PBPA0572;
OS Photobacterium profundum (Photobacterium sp. (strain SS9)).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Photobacterium.
OX NCBI_TaxID=74109;
RN [1]
RP SEQUENCE FROM N.A.
RA Vezzi A., Campanaro S., D'Angelo M., Simonato F., Vitulo N., Lauro F.,
RA Cestaro A., Malacrida G., Simonati B., Cannata N., Bartlett D.,
RA Valle G.;
RT "Genome analysis of Photobacterium profundum reveals the complexity of
RT high pressure adaptations.";
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 sigma-54 factor interaction ATP-binding
CC domain.
DR EMBL; CR378664; CAG18995.1; -.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. . .); IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR011006; CheY-like.
DR InterPro; IPR008931; FTH-Like.
DR InterPro; IPR002197; HTH_Fis.
DR InterPro; IPR002078; Sig54_interact.
DR Pfam; PF02954; HTH_8; 1.
DR Pfam; PF00158; Sigma54_activat; 1.
DR PRINTS; PR01590; HTHFIS.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR01199; HTH_fis; 1.
DR PROSITE; PS00045; SIGMA54_INTERACT_4; 1.
KW ATP-binding; Complete proteome; DNA-binding; Transcription;
KW Transcription regulation.
SQ SEQUENCE 485 AA; 54241 MW; 18D39CEBD9DBE23C CRC64;

Query Match 0.8%; Score 8; DB 2; Length 485;

```

```
Best Local Similarity 100.0%; Pred. No. 2e+02; 0; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

Qy 613 SKLTIREE 620
Db 427 SKLTIREE 434

RESULT 67
Q42989
ID Q42989 PRELIMINARY; PRT; 488 AA.
AC Q42989;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DE Beta-amylose (EC 3.2.1.2).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RC STRAIN=Tainung 67;
RP SEQUENCE FROM N.A.
RA Chen J.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; L10345; AAA33898.1; -.
DR HSSP; F16098; 1BIY.
DR Gramene; Q42989; -.
DR GO; GO:0016161; F:beta-amylose activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE 1; 1.
DR PROSITE; PS00679; BETA_AMYLASE 2; 1.
DR PROSITE; PS00136; SUBTILASE ASP; UNKNOWN 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 488 AA; 55093 MW; 30350B1AE7A8DADF CRC64;

Query Match 0.8%; Score 8; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 68
Q42990
ID Q42990 PRELIMINARY; PRT; 488 AA.
AC Q42990;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DE Beta-amylose (EC 3.2.1.2).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]

Query Match 0.8%; Score 8; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 69
Q425B2
ID Q425B2 PRELIMINARY; PRT; 488 AA.
AC Q625B2;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
DE Putative beta-amylose.
GN Name=OJ1729_E01.25;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]

SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
RT clone:OJ1729_E01.25";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AP005156; BAC83773.1; -.
DR HSSP; F16098; 1BIY.
DR GO; GO:0016161; F:beta-amylose activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001554; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE 1; 1.
```

```

DR PROSITE; P500679; BETA_AMYLASE_2; 1.
DR PROSITE; P500136; SUBTILASE ASP; UNKNOWN 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 488 AA; 55117 MW; 9E3F42A8F93EC5D6 CRC64;

Query Match      0.8%; Score 8; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 70
Q8WPP2 PRELIMINARY; PRT; 490 AA.
AC Q8WPP2;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE 6-phospho-1-fructokinase (EC 2.7.1.11).
GN Name=pfk;
OS Trypanoplasma borreli.
OC Eukaryota; Euclenozoa; Kinetoplastida; Bodonidae; Trypanoplasma.
OX NCBI_TaxID=5710;
RN [1]_SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RX MEDLINE=22168988; PubMed=12180974;
RA Lopez C., Chevallier N., Hannaert V., Rigden D.J., Michels P.A.,
RA Ramirez J.L.;
RT "Leishmania donovani phosphofructokinase: gene characterization,
RT biochemical properties and structure-modelling studies.";
RL Eur. J. Biochem. 269:3978-3989(2002).
DR EMBL; AJ310928; CAC84571.1; -.
DR GO; GO:0005945; C:6-phosphofructokinase complex; IEA.
DR GO; GO:0003872; F:6-phosphofructokinase activity; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008096; P:glycolysis; IEA.
DR Pfam; PF00365; PFK; 1.
DR PRINTS; PR00476; PHFRCTKINASE.
KW Kinase; Transferase.
SQ SEQUENCE 490 AA; 53376 MW; 62360A4071A700DF CRC64;

Query Match      0.8%; Score 8; DB 2; Length 490;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 632 VASALSHI 639
Db 287 VASALSHI 294

RESULT 71
Q6C9U7 PRELIMINARY; PRT; 492 AA.
AC Q6C9U7;
DT 25-OCT-2004 (TREMBLrel. 28, Created)
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Similar to w|NCU01449.1 Neurospora crassa NCU01449. 1 hypothetical
DE protein.
GN ORFName=YAI0D082509;
OS Yarrowia lipolytica Cl1999.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=284591;
RN [1]_SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=CL1B99;
RG Genolevures;

RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boissarie A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Teklaia F., Wesolowski-Louvel M., Westhof E., With B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissbach J.,
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]_SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=CL1B99;
RC Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBSJ databases.
DR EMBL; CR382130; CAG80753.1; -.
DR GO; GO:0003935; F:GTP cyclohydrolase II activity; IEA.
DR GO; GO:0009231; P:vitamin B2 biosynthesis; IEA.
KW Hypothetical protein.
SQ SEQUENCE 492 AA; 54208 MW; EB4A540581E29863 CRC64;

Query Match      0.8%; Score 8; DB 2; Length 492;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 879 YHALAVAT 886
Db 134 YHALAVAT 141

RESULT 72
AMVB WHEAT
ID AMVB WHEAT STANDARD; PRT; 503 AA.
AC P93594;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Beta-amylase (EC 3.2.1.2) (1,4-alpha-D-glucan maltohydrolase).
GN Name=BMV1; Synonyms=AMV1;
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]_SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Star; TISSUE=Leaf;
RC Wagner G.B., Haeger K.-P., Ziegler P.;
RT "Nucleotide sequence of a cDNA from wheat leaves encoding ubiquitous
RT beta-amylase.";
RL (er) Plant Gene Register PGR96-123.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -|- SIMILARITY: Belongs to the glycosyl hydrolase 14 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X98504; CAA67128.1; -.
CC HSP; P16098; 1BIY.
CC InterPro; IPR001554; Glyco_hydro_14.

```

```

DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETA_AMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE 1; 1.
DR PROSITE; PS00679; BETA_AMYLASE 2; 1.
DR Glycosidease; Hydrolase; Polysaccharide degradation.
FT ACT_SITE 184 184 By similarity.
FT ACT_SITE 378 378 By similarity.
SQ SEQUENCE 503 AA; 56611 MW; 7189E1533A6C0F73 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 73
GPMI_MYCGA STANDARD; PRT; 503 AA.
AC Q7NAQ5;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE 2,3-bisphosphoglycerate-independent phosphoglycerate mutase
DE (EC 5.4.2.1) (Phosphoglyceromutase) (BPG-independent PGAM) (IPGM).
GN Name=gpmi; OrderedLocNames=MYCGA5800; ORFNames=MGA_0356;
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K (low);
RX MEDLINE=22830409; PubMed=12949158; DOI=10.1099/mic.0.26427-0;
RA Papazisi L., Gorton T.S., Kutish G., Markham P.F., Browning G.F.,
RA Nguyen D.K., Swartzell S., Madan A., Mahairas G., Geary S.J.;
RT "The complete genome sequence of the avian pathogen Mycoplasma
gallisepticum strain R (low).";
RL Microbiology 149:2307-2316(2003).
CC -1- FUNCTION: Catalyzes the interconversion of 2-phosphoglycerate and
CC 3-phosphoglycerate (By similarity).
CC -1- CATALYTIC ACTIVITY: 2-phospho-D-glycerate = 3-phospho-D-glycerate.
CC -1- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Glycolysis.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SIMILARITY: Belongs to the BPG-independent phosphoglycerate mutase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE016969; AAP56930.1; -.
CC HAMAP; MF_01038; -.
CC InterPro; IPR004456; APGAM.
CC InterPro; IPR011258; IPGM N.
CC InterPro; IPR006124; Metalloenzyme.
CC InterPro; IPR005995; Pgm_bpd_ind.
CC Pfam; PF06415; IPGM N; 1.
CC Pfam; PF01676; Metalloenzyme; 1.
CC PIRSF; PIRSF001492; IPGM; 1.
CC ProDom; PD004429; Pgm_bpd_ind; 1.
CC TIGRfam; TIGR01307; pgm_bpd_ind; 1.
KW Complete proteome; Glycolysis; Isomerase; Manganese; Metal-binding.
FT ACT_SITE 60 60 Phosphoserine intermediate (By
FT similarity).
FT METAL 10 10 Manganese 2 (By similarity).

```

```

FT METAL 60 60 Manganese 2 (By similarity).
FT METAL 396 396 Manganese 1 (By similarity).
FT METAL 400 400 Manganese 1 (By similarity).
FT METAL 437 437 Manganese 2 (By similarity).
FT METAL 438 438 Manganese 2 (By similarity).
FT METAL 455 455 Manganese 1 (By similarity).
SQ SEQUENCE 503 AA; 56579 MW; 3FDD2C55BB396979 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 532 HINHIYAI 539
Db 123 HINHIYAI 130

RESULT 74
Q08335 PRELIMINARY; PRT; 503 AA.
ID Q08335;
AC Q08335;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Beta-amylase
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Secale.
OX NCBI_TaxID=4550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94151427; PubMed=8108499; DOI=10.1104/pp.102.1.315;
RA Sadowski J., Rorat T., Cooke R., Deisenay M.;
RT "Nucleotide sequence of a cDNA clone encoding ubiquitous beta-amylase
in rye (Secale cereale L.).";
RL Plant Physiol. 102:315-316(1993).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -1- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
CC
CC EMBL; Z11772; CAA77817.1; -.
CC PIR; JQ2248; JQ2248.
CC HSP; P16098; 1B1Y.
CC GO; GO:0016161; F:beta-amylase activity; IEA.
CC GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
CC GO; GO:0005975; P:carbohydrate metabolism; IEA.
CC GO; GO:0000272; P:polysaccharide catabolism; IEA.
CC InterPro; IPR001554; Glyco_hydro_14.
CC InterPro; IPR001371; Glyco_hydro_14B.
CC Pfam; PF01373; Glyco_hydro_14; 1.
CC PRINTS; PR00750; BETA_AMYLASE.
CC PRINTS; PR00842; GLYDLAS14B.
CC PROSITE; PS00506; BETA_AMYLASE 1; 1.
CC PROSITE; PS00679; BETA_AMYLASE 2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 503 AA; 56759 MW; D788B79A8B8D9523 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811
Db 55 GLVEGKGP 62

RESULT 75
Q6SNP7 PRELIMINARY; PRT; 517 AA.
ID Q6SNP7;
AC Q6SNP7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)

```

DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)  
 DE Beta-amylase 1 (Fragment).  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Clark S.E., Hayes P.M., Henson C.A.;  
 RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in  
 CC polysaccharides so as to remove successive maltose units from the  
 CC non-reducing ends of the chains.  
 CC -1- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.  
 DR EMBL; AY454398; RAR10251.1; -  
 DR HSSP; P16098; 1BY.  
 DR GO; GO:0016161; F:beta-amylase activity; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO; GO:000272; P:polysaccharide catabolism; IEA.  
 DR InterPro; IPR001554; Glyco\_hydro\_14.  
 DR InterPro; IPR001371; Glyco\_hydro\_14B.  
 DR Pfam; PF01373; Glyco\_hydro\_14; 1.  
 DR PRINTS; PR00750; BETAMYLASE.  
 DR PRINTS; PR00842; GLHYDLASE14B.  
 DR PROSITE; PS00506; BETA\_AMYLASE\_1; 1.  
 DR PROSITE; PS00679; BETA\_AMYLASE\_2; 1.  
 DR Carbohydrate metabolism; Glycosidase; Hydrolase;  
 KW Polysaccharide degradation.  
 FT NON\_TER 1  
 FT NON\_TER 517 517  
 SQ SEQUENCE 517 AA; 57635 MW; F41B4950559871BB CRC64;  
  
 Query Match 0.8%; Score 8; DB 2; Length 517;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 804 GLVECKGP 811  
 Db |||||  
 43 GLVECKGP 50  
  
 RESULT 76  
 Q7MD66 PRELIMINARY; PRT; 528 AA.  
 ID Q7MD66  
 AC Q7MD66;  
 DT 01-MAR-2004 (TRENBLrel. 26, Created)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Putative N-acyl-D-glutamate deacylase protein.  
 GN OrderedLocNames=VVAL170;  
 OS Vibrio vulnificus (strain VJ016).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=196600;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA PubMed=14656965; DOI=10.1101/gr.1295503;  
 RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,  
 RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,  
 RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-P.;  
 RT "Comparative genome analysis of Vibrio vulnificus, a marine  
 RT pathogen.";  
 RL Genome Res. 13:2577-2587(2003).  
 DR EMBL; AP005349; BAC97196.1; -  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR006680; Amidohydro\_1  
 DR InterPro; IPR011059; Metallo\_hydrolase.  
 DR Pfam; PF01979; Amidohydro\_1; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 528 AA; 58175 MW; 584BF68E20C68097 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 528;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 689 EKILKIE 696  
 Db |||||  
 186 EKILKIE 193  
  
 RESULT 77  
 Q9FUK6 PRELIMINARY; PRT; 533 AA.  
 ID Q9FUK6;  
 AC Q9FUK6;  
 DT 01-MAR-2001 (TRENBLrel. 16, Created)  
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Beta-amylase (Fragment).  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ma Y., Evans E.D., Logue S.J., Langridge P.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in  
 CC polysaccharides so as to remove successive maltose units from the  
 CC non-reducing ends of the chains.  
 CC -1- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.  
 DR EMBL; AF300800; AAG25638.1; -  
 DR HSSP; P16098; 1BY.  
 DR GO; GO:0016161; F:beta-amylase activity; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO; GO:000272; P:polysaccharide catabolism; IEA.  
 DR InterPro; IPR001554; Glyco\_hydro\_14.  
 DR InterPro; IPR001371; Glyco\_hydro\_14B.  
 DR Pfam; PF01373; Glyco\_hydro\_14; 1.  
 DR PRINTS; PR00750; BETAMYLASE.  
 DR PRINTS; PR00842; GLHYDLASE14B.  
 DR PROSITE; PS00506; BETA\_AMYLASE\_1; 1.  
 DR PROSITE; PS00679; BETA\_AMYLASE\_2; 1.  
 KW Carbohydrate metabolism; Glycosidase; Hydrolase;  
 KW Polysaccharide degradation.  
 FT NON\_TER 1  
 FT NON\_TER 533 533  
 SQ SEQUENCE 533 AA; 59312 MW; C0B9780B0F13F48C CRC64;  
  
 Query Match 0.8%; Score 8; DB 2; Length 533;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 804 GLVECKGP 811  
 Db |||||  
 53 GLVECKGP 60  
  
 RESULT 78  
 Q9FUK7 PRELIMINARY; PRT; 533 AA.  
 ID Q9FUK7;  
 AC Q9FUK7;  
 DT 01-MAR-2001 (TRENBLrel. 16, Created)  
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Beta-amylase (Fragment).  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.



```

RA Ma Y., Evans E.D., Loque S.J., Langridge P.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBSJ databases.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -|- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AF300799; AAG25637.1; -.
DR HSSP; P16098; 1BY.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001554; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR PRINTS; PR00750; BETAAMYLASE.
DR PRINTS; PR00842; GLHYDLASE14B.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ NON_TER 1
SQ SEQUENCE 533 AA; 59387 MW; 793031ED53FFACB5 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 533;
Best Local Similarity 100.0%; Pred. No. 2.28+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
|||||
DB 53 GLVEGKGP 60

RESULT 79
AMVB HORVU STANDARD; PRT; 535 AA.
ID AMVB HORVU STANDARD; PRT; 535 AA.
AC P16098;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Beta-amylase (EC 3.2.1.2) (1,4-alpha-D-glucan maltohydrolase).
GN Name=BMV1; Synonyms=AMVB;
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Endosperm;
RX MEDLINE=88082785; PubMed=2446870;
RA Kreis M., Williamson M., Buxton B., Pywell J., Hejgaard J.,
RA Svendsen I.;
RT "Primary structure and differential expression of beta-amylase in
RT normal and mutant barleys";
RL Eur. J. Biochem. 169:517-525 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Haruna Two-rows;
RX MEDLINE=96068953; PubMed=8534999;
RA Yoshigi N., Okada Y., Sahara H., Tamaki T.;
RT "A structural gene encoding beta-amylase of barley.";
RL Biosci. Biotechnol. Biochem. 59:1991-1993 (1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Haruna Two-rows;
RX MEDLINE=94245663; PubMed=8188635;
RA Yoshigi N., Okada Y., Sahara H., Koshino S.;
RT "PCR cloning and sequencing of the beta-amylase cDNA from barley.";
RL J. Biochem. 115:47-51 (1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 5-504.
RX MEDLINE=99141375; PubMed=9918723; DOI=10.1006/jmbi.1998.2379;

```

```

RA Mikami B., Yoon H.-J., Yoshigi N.;
RT "The crystal structure of the sevenfold mutant of barley beta-amylase
RT with increased thermostability at 2.5 A resolution.";
RL J. Mol. Biol. 285:1235-1243 (1999).
CC -|- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -|- SUBUNIT: Monomer.
CC -|- SIMILARITY: Belongs to the glycosyl hydrolase 14 family.
CC -|- THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/ebis/
CC or send an email to license@ebi.ac.uk).
DR EMBL; X52321; CAA36556.1; -.
DR EMBL; D49999; BAA08741.1; -.
DR EMBL; D21349; BAA04815.1; -.
DR PIR; S00222; S00222.
DR PDB; 1BY; X-ray; A=5-504.
DR InterPro; IPR001554; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAAMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW 3D-structure; Direct protein sequencing; Glycosidase; Hydrolase;
KW Polysaccharide degradation; Repeat.
FT ACT_SITE 184 By similarity.
FT ACT_SITE 378 By similarity.
FT DOMAIN 489 532 4 X 11 AA tandem repeats.
FT REPEAT 489 499 1.
FT REPEAT 500 510 2.
FT REPEAT 511 521 3.
FT REPEAT 522 532 4 (approximate).
FT CONFLICT 233 233 V -> A (in Ref. 2 and 3).
FT CONFLICT 347 347 L -> S (in Ref. 2 and 3).
FT CONFLICT 527 527 I -> M (in Ref. 2 and 3).
FT HELIX 6 8
FT STRAND 11 15
FT TURN 18 19
FT TURN 23 24
FT TURN 29 29
FT HELIX 30 42
FT TURN 43 44
FT STRAND 47 53
FT TURN 54 56
FT HELIX 57 60
FT TURN 62 63
FT TURN 68 80
FT TURN 81 81
FT STRAND 83 89
FT STRAND 93 93
FT TURN 97 98
FT STRAND 103 103
FT HELIX 107 115
FT HELIX 117 119
FT STRAND 120 122
FT TURN 124 125
FT STRAND 128 133
FT HELIX 135 137
FT TURN 138 139
FT TURN 142 144
FT HELIX 148 169
FT TURN 170 170
FT STRAND 172 177
FT HELIX 181 183
FT STRAND 184 184
FT HELIX 192 194
FT TURN 195 195

```

FT TURN 198 199  
 FT TURN 209 221  
 FT TURN 222 223  
 FT TURN 225 226  
 FT TURN 231 232  
 FT TURN 236 237  
 FT TURN 240 242  
 FT TURN 244 246  
 FT TURN 248 249  
 FT TURN 251 253  
 FT TURN 255 283  
 FT TURN 284 285  
 FT STRAND 289 294  
 FT TURN 299 300  
 FT TURN 307 312  
 FT TURN 313 313  
 FT TURN 322 323  
 FT TURN 324 330  
 FT TURN 331 334  
 FT STRAND 336 339  
 FT TURN 342 343  
 FT TURN 346 348  
 FT TURN 351 354  
 FT TURN 357 370  
 FT TURN 371 372  
 FT STRAND 375 377  
 FT TURN 386 396  
 FT TURN 398 399  
 FT TURN 403 404  
 FT STRAND 413 416  
 FT TURN 421 423  
 FT TURN 426 439  
 FT TURN 440 442  
 FT TURN 464 467  
 FT TURN 468 470  
 SQ SEQUENCE 535 AA; 59647 MW; FFDA4CED53E9A89E CRC64;

Query Match 0.8%; Score 8; DB 1; Length 535;

Best Local Similarity 100.0%; Pred.No.2.2e+02; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811  
 |||||  
 Db 55 GLVEGKGP 62

RESULT 80  
 ID P82993 PRELIMINARY; PRT; 535 AA.  
 AC P82993;  
 DT 01-OCT-2001 (TREMELrel. 18, Created)  
 DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)  
 DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)  
 DE Beta-amylase precursor (EC 3.2.1.2) (Beta-Amy1) (1,4-alpha-D-glucan  
 maltohydrolase).  
 DE Name=BMV1; Synonyms=AMYB;  
 OS Hordeum spontaneum (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=77009;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND FUNCTION.  
 RC STRAIN=NPGS P1296897; TISSUE=Endosperm;  
 RX MEDLINE=98289114; PubMed=9625721; DOI=10.1104/pp.117.2.679;  
 RA Erkkila M.J., Leah R., Ahokas H., Cameron-Mills V.;  
 RT "Allele-dependent barley grain beta-amylase activity.";  
 RL Plant Physiol. 117:679-685(1998).  
 RN [2]  
 RP PARTIAL SEQUENCE, FUNCTION, TISSUE SPECIFICITY, ACETYLATION, AND  
 RP VARIANTS ASP-165; SER-254 AND GLN-472.  
 RC STRAIN=cv. CPI77146-33; TISSUE=Seed;  
 RA Eglinton J.K., Lahnstein J., Shirley N., Evans D.E.;

RT "The structural basis for functional differences between allelic forms  
 of barley (Hordeum vulgare) beta-amylase.";  
 RL Eur. J. Biochem. 0:0-0(2001).  
 CC -|- FUNCTION: Catalyzes the liberation of maltose from 1,4-alpha-D  
 glucans.  
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in  
 polysaccharides so as to remove successive maltose units from the  
 non-reducing ends of the chains.  
 CC -|- SUBUNIT: Monomer (By similarity).  
 CC -|- TISSUE SPECIFICITY: Endosperm.  
 CC -|- POLYMORPHISM: There are at least three alleles; SD2H; SD1 and  
 SD2L. The sequence of SD2H is shown here.  
 CC -|- MISCELLANEOUS: The three alleles show different thermostabilities  
 and different affinities for soluble starch.  
 CC -|- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.  
 DR EMBL; AF061204; AAC67246.1; -.  
 DR HSP; F16098; IBY.  
 DR GO; GO:0016161; F:beta-amylase activity; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO; GO:0000272; P:polysaccharide catabolism; IEA.  
 DR InterPro; IPR001554; Glyco\_hydro\_14.  
 DR InterPro; IPR001371; Glyco\_hydro\_14; 1.  
 DR Pfam; PF01373; Glyco\_hydro\_14; 1.  
 DR PRINTS; PR00750; BETAAMYLASE.  
 DR PRINTS; PR00842; GLHYDLASE14B.  
 DR PROSITE; PS00506; BETA\_AMYLASE\_1; 1.  
 DR PROSITE; PS00679; BETA\_AMYLASE\_2; 1.  
 DR Acetylation; Glycosidase; Hydrolase; Polymorphism;  
 KW Polysaccharide degradation; Repeat.  
 FT PROPEP 1 2  
 FT CHAIN 3 489 Removed in mature form.  
 FT PROPEP 3 489 Beta-amylase.  
 FT PROPEP 490 535 Removed in mature form.  
 FT MOD\_RES 3 3 N-acetylvaline.  
 FT ACT\_SITE 99 99 By similarity.  
 FT ACT\_SITE 184 184 By similarity.  
 FT DOMAIN 489 532 4 X 11 AA tandem repeats.  
 FT REPEAT 489 499 1.  
 FT REPEAT 500 510 2.  
 FT REPEAT 511 521 3.  
 FT REPEAT 522 532 4 (Approximate).  
 FT VARIANT 165 165 E -> D.  
 FT VARIANT 254 254 T -> S.  
 FT VARIANT 472 472 K -> Q.  
 FT SEQUENCE 535 AA; 59639 MW; 0B265EAEF061C9B CRC64;

Query Match 0.8%; Score 8; DB 2; Length 535;

Best Local Similarity 100.0%; Pred.No.2.2e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811  
 |||||  
 Db 55 GLVEGKGP 62

RESULT 81  
 ID Q84T19 PRELIMINARY; PRT; 535 AA.  
 AC Q84T19;  
 DT 01-JUN-2003 (TREMELrel. 24, Created)  
 DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Endosperm-specific beta-amylase 1.  
 GN Name=Bamy1;  
 OS Hordeum vulgare var. distichum (Two-rowed barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=112509;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Clark S.E., Hayes P.M., Henson C.A.;  
 RT "Effects of single nucleotide polymorphisms in beta-amylase alleles

```

RT from barley on functional properties of the enzymes." ;
RL Plant Physiol. Biochem. 41:798-804(2003).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AF414082; AAO67356.1; -.
DR HSSP; P16098; 1BIY.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001354; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAMYLASE.
DR PRINTS; PR00842; GLHYDLASE14B.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 535 AA; 59639 MW; 338FDCFFC66A51C CRC64;

Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62

RESULT 82
Q84T20 ID Q84T20 PRELIMINARY; PRT; 535 AA.
AC Q84T20;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-OCT-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endosperm-specific beta-amylase 1.
GN Name=bamyl;
OS Hordeum vulgare var. distichum (Two-rowed barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=112509;
RN [1]
RP SEQUENCE FROM N.A.
RA Clark S.E., Hayes P.M., Henson C.A.;
RA "Effects of single nucleotide polymorphisms in beta-amylase alleles
RT from barley on functional properties of the enzymes." ;
RL Plant Physiol. Biochem. 41:798-804(2003).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AF414081; AAO67355.1; -.
DR HSSP; P16098; 1BIY.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001354; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAMYLASE.
DR PRINTS; PR00842; GLHYDLASE14B.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 535 AA; 59572 MW; 4653050B0F05F0A7 CRC64;

from barley on functional properties of the enzymes." ;
RL Plant Physiol. Biochem. 41:798-804(2003).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AF414082; AAO67356.1; -.
DR HSSP; P16098; 1BIY.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001354; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAMYLASE.
DR PRINTS; PR00842; GLHYDLASE14B.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 535 AA; 59572 MW; 4653050B0F05F0A7 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62

RESULT 83
Q9AVJ8 ID Q9AVJ8 PRELIMINARY; PRT; 535 AA.
AC Q9AVJ8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sdi beta-amylase.
GN Name=Bmy1;
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RA Li C., Langridge P., Zhang X., Eckstein P.E., Rosenagel B.G.,
RA Lance R.C., Lefol E.B., Lu M., Harvey B.L., Scoles G.J.;
RA Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RL -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -!- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.
DR EMBL; AB048949; BAB39391.1; -.
DR HSSP; P16098; 1BIY.
DR GO; GO:0016161; F:beta-amylase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001354; Glyco_hydro_14.
DR InterPro; IPR001371; Glyco_hydro_14B.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAMYLASE.
DR PRINTS; PR00842; GLHYDLASE14B.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Carbohydrate metabolism; Glycosidase; Hydrolase;
KW Polysaccharide degradation.
SQ SEQUENCE 535 AA; 59572 MW; 4653050B0F05F0A7 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62

RESULT 84
Q9FSI3 ID Q9FSI3 PRELIMINARY; PRT; 535 AA.
AC Q9FSI3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Beta-amylase.
GN Name=Bmy1;
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;

```

RN SEQUENCE FROM N.A.  
 RP MEDLINE=21416776; PubMed=11525070;  
 RA Erkkila M.J., Hannu A.;  
 RT "Special barley beta-amylase allele in a Finnish landrace line HNS2  
 RT with high grain enzyme activity.";  
 RL Hereditas 134:91-95(2001).  
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in  
 CC polysaccharides so as to remove successive maltose units from the  
 CC non-reducing ends of the chains.  
 CC -I- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.  
 DR EMBL; AJ301645; CAC16789.1; -;  
 DR HSSP; P16098; 1B1Y.  
 DR GO; GO:0016161; F:beta-amylase activity; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO; GO:0000272; P:polysaccharide catabolism; IEA.  
 DR InterPro; IPR001554; Glyco\_hydro\_14.  
 DR InterPro; IPR001371; Glyco\_hydro\_14B.  
 DR Pfam; PF01373; Glyco\_hydro\_14; 1.  
 DR PRINTS; PR00750; GLHYDLASE14B.  
 DR PRINTS; PR00842; GLHYDLASE14B.  
 DR PROSITE; PS00506; BETA\_AMYLASE\_1; 1.  
 DR PROSITE; PS00679; BETA\_AMYLASE\_2; 1.  
 DR Carbohydrate metabolism; Glycosidase; Hydrolase;  
 KW Polysaccharide degradation.  
 SQ SEQUENCE 535 AA; 59538 MW; 1F30B0865463D6E5 CRC64;  
 Query Match 0.8%; Score 8; DB 2; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 804 GLVEGKGP 811  
 Db |||||  
 55 GLVEGKGP 62  
 RESULT 85  
 Q9SBH7 PRELIMINARY; PRT; 535 AA.  
 AC Q9SBH7  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Beta-amylase [EC 3.2.1.2].  
 GN Name-beta-amyl;  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
 OC Triticeae; Hordeum.  
 OC NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98289114; PubMed=9625721; DOI=10.1104/pp.117.2.679;  
 RA Erkkila M.J., Leah R., Ahokas H., Cameron-Mills V.;  
 RT "Allele-dependent barley grain beta-amylase activity.";  
 RL Plant Physiol. 117:679-695(1998).  
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in  
 CC polysaccharides so as to remove successive maltose units from the  
 CC non-reducing ends of the chains.  
 CC -I- SIMILARITY: Belongs to family 14 of glycosyl hydrolases.  
 DR EMBL; AF061203; AAC67245.1; -;  
 DR HSSP; P16098; 1B1Y.  
 DR GO; GO:0016161; F:beta-amylase activity; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO; GO:0000272; P:polysaccharide catabolism; IEA.  
 DR InterPro; IPR001554; Glyco\_hydro\_14.  
 DR InterPro; IPR001371; Glyco\_hydro\_14B.  
 DR Pfam; PF01373; Glyco\_hydro\_14; 1.  
 DR PRINTS; PR00750; BETAAMYLASE.  
 DR PRINTS; PR00842; GLHYDLASE14B.  
 DR PROSITE; PS00506; BETA\_AMYLASE\_1; 1.

DR PROSITE; PS00679; BETA\_AMYLASE\_2; 1.  
 KW Carbohydrate metabolism; Glycosidase; Hydrolase;  
 KW Polysaccharide degradation.  
 SQ SEQUENCE 535 AA; 59665 MW; FACE4CED53E9A89E CRC64;  
 Query Match 0.8%; Score 8; DB 2; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 804 GLVEGKGP 811  
 Db |||||  
 55 GLVEGKGP 62  
 RESULT 86  
 Q7X3X6 PRELIMINARY; PRT; 543 AA.  
 ID Q7X3X6  
 AC Q7X3X6  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Putative tomatinase Toma.  
 GN Name-toma;  
 OS Clavibacter michiganensis subsp. michiganensis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Micrococcineae; Microbacteriaceae; Clavibacter.  
 OC NCBI\_TaxID=33013;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCPB382;  
 RA Gartenmann K.-H., Graefen I., Zellermann E.-M., Burger A.,  
 RA Eichenlaub R.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF393183; AAP57293.1; -;  
 DR HSSP; P23360; 111W.  
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR001000; Glyco\_hydro\_10.  
 DR Pfam; PF00331; Glyco\_hydro\_10; 1.  
 DR SMART; SM00633; Glyco\_10; 1.  
 DR SEQUENCE 543 AA; 58175 MW; 8EC284F8A521AED CRC64;  
 Query Match 0.8%; Score 8; DB 2; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 799 LDLLTGLV 806  
 Db |||||  
 227 LDLLTGLV 234  
 RESULT 87  
 Q6V1X2 PRELIMINARY; PRT; 546 AA.  
 ID Q6V1X2  
 AC Q6V1X2  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE SFRS protein kinase 2 isoform c.  
 GN Name=SRPK2;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Sha J.H., Zhou Z.M., Li J.M.;  
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
 CC -I- SIMILARITY: Belongs to the Ser/Thr protein kinase family.  
 DR EMBL; AY354201; AAQ63886.1; -;  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.

```

DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR011009; Kinase like.
DR InterPro; IPR000719; Ser kinase.
DR InterPro; IPR002290; Ser Thr kinase.
DR InterPro; IPR008271; Ser Thr_kin_AS.
DR InterPro; IPR001245; Tyros_kinase.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TK; 1.
DR SMART; SM00219; TyRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 546 AA; 61049 MW; 5F1DB6805A6CD79C CRC64;

Query Match 0.8%; Score 8; DB 2; Length 546;
Best Local Similarity 100.0%; Pred.No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSE 951
DB 5 SEKSSSE 12

RESULT 88
Q6WKW9 PRELIMINARY; PRT; 554 AA.
AC Q6WKW9;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DE HMg-box protein HMG2L1.
GN Name=HMG2L1;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2275817; PubMed=12975653;
RA Yamada M., Ohkawara B., Ichimura N., Hyodo-Miura J., Urushiyama S.,
RA Shirakabe K., Shibuya H.;
RT "Negative regulation of Wnt signalling by HMG2L1, a novel NLK-binding
protein.";
RL Genes Cells 8:677-684(2003).
DR EMBL; AY281248; RAP92714.1; -.
DR HSP; P07155; IAB.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; F:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000135; Highmobility_12.
DR InterPro; IPR000910; HMg_box.
DR Pfam; PF05055; HMg_box; 1.
DR PRINTS; PR00886; HIGHMOBILITY12.
DR SMART; SM00398; HMg; 1.
DR PROSITE; PS01118; HMg_BOX_2; 1.
SQ SEQUENCE 554 AA; 60909 MW; A2DF64DCA5532P36 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 554;
Best Local Similarity 100.0%; Pred.No. 2.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 946 KSSSSSE 953
DB 430 KSSSSSE 437

RESULT 89

```

```

Q38914 PRELIMINARY; PRT; 555 AA.
ID Q38914;
AC Q38914;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE CKC.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsia.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U44028; AAA86281.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; F:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001471; TP_ERF.
DR PRINTS; PR00367; ETRSPSELMNT.
DR ProDom; PD001423; TF_ERF; 2.
DR SMART; SM00380; AP2_2.
SQ SEQUENCE 555 AA; 61741 MW; A6275D2F082B562D CRC64;

Query Match 0.8%; Score 8; DB 2; Length 555;
Best Local Similarity 100.0%; Pred.No. 2.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVLTL 12
DB 536 LKTVLTL 543

RESULT 90
Q42462 PRELIMINARY; PRT; 555 AA.
ID Q42462;
AC Q42462;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE ANT (Ovule development protein aintegumenta).
GN Name=AINTEGUMENTA; Synonym=AT4937750; At4G37750;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsia.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Landsberg erecta; TISSUE=Flower;
RX MEDLINE=963151414; PubMed=8742707;
RA Elliott R.C., Betzner A.S., Huttner E., Oakes M.P., Tucker W.Q.,
RA Gerentes D., Perez P., Smyth D.R.;
RT "AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic
roles in ovule development and floral organ growth.";
RL Plant Cell 8:155-168(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Robben J., Grymonprez B., Volckaert G., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Klucher K.J., Chow H., Reiser L., Fischer R.L.;
RT "The AINTEGUMENTA Gene of Arabidopsis Required for Ovule and Female
Gametophyte Development is Related to the Floral Homeotic Gene
APETALA2.";
RT APETALA2.";

```

Plant Cell 8:0-0(1996).  
 RA [5]  
 RP SEQUENCE FROM N.A.  
 RA Klucher K.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RA Bewan M., Robben J., Grymonprez B., Volckaert G., Bancroft I.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U41339; AAB17364.1; -;  
 DR EMBL; AL161592; CAB80440.1; -;  
 DR EMBL; U40256; AAA91040.1; -;  
 DR EMBL; AY080706; AAL85024.1; -;  
 DR EMBL; AY117207; AAM51282.1; -;  
 DR EMBL; AL035709; CAB38923.1; -;  
 DR PIR; S71365; S71365.  
 DR TRANSFAC; T02639; -;  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001471; TF\_ERF.  
 DR PRINTS; PR00367; ETRSPLEEMNT.  
 DR ProDom; PD001423; TF\_ERF\_2.  
 DR SMART; SM00380; AP2; 2.  
 SQ SEQUENCE 555 AA; 61725 MW; BD275D2F082B519B CRC64;  
 Query Match 0.8%; Score 8; DB 2; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02; Mismatches 0; Indels 0; Gaps 0;  
 Matches 8; Conservative 0;  
 QY 5 LKTVALTL 12  
 Db 536 LKTVALTL 543  
 RESULT 91  
 Q9P4A3 PRELIMINARY; PRT; 594 AA.  
 AC Q9P4A3;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Camptothecin resistance conferring protein rcaa.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; Emericella.  
 OX NCBI\_TaxID=162425;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21258791; PubMed=11361337; DOI=10.1007/s004380000411;  
 RA Bruschi G.C., de Souza C.C., Fagundes M.R., Dani M.A., Goldman M.H.,

RA Morris N.R., Liu L., Goldman G.H.;  
 RT "Sensitivity to camptothecin in Aspergillus nidulans identifies a  
 RT novel gene, scaA+, related to the cellular DNA damage response.";  
 RL Mol. Genet. Genomics 265:264-275(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Bruschi G.B.C.M., de Souza C.C., Dani M.A.C., Terenzi M.F.,  
 RA Goldman M.H.S., Liu L., Morris N.R., Goldman G.H.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF228504; AAF81094.1; -;  
 SQ SEQUENCE 594 AA; 67513 MW; 19C4BE9E3DBD4505 CRC64;  
 Query Match 0.8%; Score 8; DB 2; Length 594;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 891 EGLDIKTI 898  
 Db 33 EGLDIKTI 40  
 RESULT 92  
 Q7VQZ4 PRELIMINARY; PRT; 625 AA.  
 ID Q7VQZ4;  
 AC Q7VQZ4;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Protease IV, a signal peptide peptidase (EC 3.4.21.-).  
 GN Name=sppA; OrderedLocusNames=Bfl136;  
 OS Candidatus Blochmannia floridanus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; ant endosymbionts; Candidatus Blochmannia.  
 OX NCBI\_TaxID=203907;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22784745; PubMed=12886019; DOI=10.1073/pnas.1533499100;  
 RA Gil R., Silva F.J., Zientz E., Delmotte F., Gonzalez-Candelas F.,  
 RA Latorre A., Rausell C., Kamerbeek J., Gadau J., Hoelldobler B.,  
 RA van Ham R.C.H.J., Gross R., Moya A.;  
 RT "The genome sequence of Blochmannia floridanus: comparative analysis  
 RT of reduced genomes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:9388-9393(2003).  
 DR EMBL; BX248585; CAD83498.1; -;  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0009003; P:signal peptidase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006465; P:signal peptide processing; IEA.  
 DR InterPro; IPR002142; Peptidase\_S49.  
 DR InterPro; IPR004635; Pept\_S49\_SppA.  
 DR InterPro; IPR004634; Pept\_S49\_SppA67.  
 DR Pfam; PF01343; Peptidase\_S49; 2.  
 DR PIRSF; PIRSF001217; Protease\_4\_SppA; 1.  
 DR ProDom; PD002897; Peptidase\_S49; 1.  
 DR TIGRFAMs; TIGR00705; SppA\_67K; 1.  
 DR TIGRFAMs; TIGR00706; SppA\_dom; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 625 AA; 71013 MW; D3D4DB1753491AF6 CRC64;  
 Query Match 0.8%; Score 8; DB 2; Length 625;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 843 KSGKLIYA 850  
 Db 141 KSGKLIYA 148  
 RESULT 93  
 Q9FFV8 PRELIMINARY; PRT; 642 AA.  
 ID Q9FFV8  
 AC Q9FFV8;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)



```
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Similarity to quinine nucleotide exchange factor (Hypothetical protein
DE At5g38640; MBI18.19) (At5g38640).
GN Name=At5g38640;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT P1 clones."
RL DNA Res. 4:215-230(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Jones T., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB005231; BAB10154.1; -
DR EMBL; AY042523; AAL32601.1; -
DR EMBL; BT008789; AAP68227.1; -
DR GO; GO:0005851; C:eukaryotic translation initiation factor 2B. . .; IEA.
DR GO; GO:0005525; F:GTP binding; IEA.
DR GO; GO:0003743; P:translational initiation factor activity; IEA.
DR GO; GO:0006413; P:translational initiation; IEA.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR000849; IF-2B.
DR Pfam; PF01008; IF-2B; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; UNKNOWN 1.
KW Hypothetical protein.
SQ SEQUENCE 642 AA; 69495 MW; BBB6507994127893 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 642;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LTTVSVVT 19
Db 149 LTTVSVVT 156

RESULT 94
Q6ZPU7 PRELIMINARY; PRT; 649 AA.
AC Q6ZPU7;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DE MKAA1286 protein (fragment).
GN Name=MKAA1286;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic tail;
RX PubMed=14621295;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA Saka Y., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT III. the complete nucleotide sequences of 500 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries."
RL DNA Res. 10:167-180(2003).
DR EMBL; AK129322; BAC98132.1; -
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR000313; PWMP.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00855; PWMP; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00293; PWMP; 1.
DR PROSITE; PS50014; BROMODOMAIN 2; 1.
DR PROSITE; PS50812; PWMP; 1.
FT NON TER 1
SQ SEQUENCE 649 AA; 73312 MW; 45F3B0C0EDACF64 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 649;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 122 LSLNKTVP 129
Db 295 LSLNKTVP 302

RESULT 95
Q6MW74 PRELIMINARY; PRT; 662 AA.
AC Q6MW74;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DE B1340F09.9 protein.
GN Name=B1340F09.9;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12447439; DOI=10.1038/nature01183;
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA Weng Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA Han B.;
RT "Sequence and analysis of rice chromosome 4."
RL Nature 420:316-320(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lu G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
```

RL Submitted (NOV-2003) to the EMBL/GenBank/DDBJ databases.  
 CC -1- SIMILARITY: Belongs to the Ser/Thr protein kinase family.  
 DR EMBL; BX842608; CAE76071.1; -.  
 DR GO; GO:0005509; F:ATP binding; IEA.  
 DR GO; GO:0005524; F:calcium ion binding; IEA.  
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.  
 DR GO; GO:0004674; F:transferase activity; IEA.  
 DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.  
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR GO; GO:0006468; P:transferase activity; IEA.  
 DR InterPro; IPR011061; Antihaemostatic.  
 DR InterPro; IPR001881; EGF Ca.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR011009; Kinase like.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF07645; EGF CA; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00179; EGF CA; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00219; TyKc; 1.  
 DR PROSITE; PS01187; EGF CA; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN\_1.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; EGF-like domain; Kinase; Serine/threonine-protein kinase;  
 KW Transferase.  
 SQ SEQUENCE 662 AA; 72871 MW; 86E15B4296F87304 CRC64;  
  
 Query Match 0.8%; Score 8; DB 2; Length 662;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 630 LEVASALS 637  
 DB 437 LEVASALS 444  
  
 RESULT 96  
 Q7XP38 PRELIMINARY; PRT; 673 AA.  
 AC Q7XP38;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE OSUNBA0027H09.1 protein.  
 GN Name=OSUNBA0027H09.1;  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=12447439; DOI=10.1038/nature01183;  
 RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,  
 RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,  
 RA Weng Q., Zhang L., Lu Y., Mu J., Zhang L.S., Yu Z., Fan D.,  
 RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,  
 RA Wu M., Zhang R., Zhou B., Chen L., Jin Z., Wang R., Yin H.,  
 RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,  
 RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,  
 RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,  
 RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,  
 RA Han B.;  
 RT "Sequence and analysis of rice chromosome 4";  
 RL Nature 420:316-320(2002).  
 CC -1- SIMILARITY: Belongs to the Ser/Thr protein kinase family.  
 DR EMBL; AL662964; CAE03801.2; -.  
 DR HSPF; P01130; 1H28.  
 DR Gramine; Q7XP38; -.

DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.  
 DR GO; GO:0004674; F:transferase activity; IEA.  
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR InterPro; IPR011061; Antihaemostatic.  
 DR InterPro; IPR001881; EGF Ca.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR011009; Kinase like.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF07645; EGF CA; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00179; EGF CA; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00219; TyKc; 1.  
 DR PROSITE; PS01187; EGF CA; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN\_1.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; EGF-like domain; Kinase; Serine/threonine-protein kinase;  
 KW Transferase.  
 SQ SEQUENCE 673 AA; 74034 MW; 07BEFBA756CAAB9D CRC64;  
  
 Query Match 0.8%; Score 8; DB 2; Length 673;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 630 LEVASALS 637  
 DB 437 LEVASALS 444  
  
 RESULT 97  
 Q7Z5E6 PRELIMINARY; PRT; 681 AA.  
 AC Q7Z5E6;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Medulloblastoma antigen MU-MB-20.201 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Medulloblastoma;  
 RA Behrends U., Mautner J.M., Muller-Wehrich S.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Medulloblastoma;  
 RA Behrends U., Roesler S., Mautner J.M.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY131226; AAN07912.1; -.  
 FT NON TER 1  
 SQ SEQUENCE 681 AA; 79727 MW; C6598C3FD2D1425C CRC64;  
  
 Query Match 0.8%; Score 8; DB 2; Length 681;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 793 KXLEKELD 800  
 DB 472 KXLEKELD 479  
  
 RESULT 98

```

O54781
ID O54781 PRELIMINARY; PRT; 681 AA.
AC O54781
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE SRPK2.
GN Name=SRpk2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroyanagi N., Onogi H., Wakabayashi T., Hagiwara M.;
RT "Novel SR-protein-specific kinase, SRPK2, disassembles nuclear
RT speckles.";
RL Biochem. Biophys. Res. Commun. 0:0-0(1998).
CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
DR EMBL; AB006036; BAA24055.1; -.
DR FIC; JCS929; JCS929.
DR HSSP; Q03656; IHOW.
DR MGD; MGI:1201408; SrpK2.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; F:protein amino acid phosphorylation; IEA.
DR InterPro; IPR011009; Kinase like.
DR InterPro; IPR000719; Prot kinase.
DR Pfam; PF00069; Pkinase; I.
DR ProDom; PD000001; Prot_kinase; 2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 681 AA; 76740 MW; 1377ADF95C48BC4E CRC64;

Query Match 0.8%; Score 8; DB 2; Length 681;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSE 951
Db |||||
5 SEKSSSSE 12

RESULT 99
Q8VCD9 PRELIMINARY; PRT; 682 AA.
ID Q8VCD9
AC Q8VCD9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Serine/arginine-rich protein specific kinase 2.
GN Name=SRpk2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

```

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
RA Fahney J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalley D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
DR EMBL; BC020178; AAH20178.1; -.
DR PIR; JCS929; JCS929.
DR HSSP; Q03656; IHOW.
DR MGD; MGI:1201408; SrpK2.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; F:protein amino acid phosphorylation; IEA.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 682 AA; 76884 MW; F519C93A00AC1B9E CRC64;

Query Match 0.8%; Score 8; DB 2; Length 682;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSE 951
Db |||||
5 SEKSSSSE 12

RESULT 100
P78362 PRELIMINARY; PRT; 686 AA.
ID P78362
AC P78362;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Serine kinase SRPK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98139536; PubMed=9472028; DOI=10.1083/jcb.140.4.737;
RA Wang H.Y., Lin W., Dyck J.A., Yeakley J.M., Songyang Z., Cantley L.C.,
RA Fu X.D.;
RT "SRPK2, a differentially expressed SR protein-specific kinase involved
RT in mediating the interaction and localization of pre-mRNA splicing
RT factors in mammalian cells.";
RL J. Cell Biol. 140:737-750(1998).
CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
DR EMBL; U88666; AAC05299.1; -.
DR HSSP; Q03656; IHOW.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0004672; F:protein kinase activity; TAS.
DR GO; GO:0008360; F:RNA splicing; TAS.
DR GO; GO:0000245; P:spliceosome assembly; TAS.
DR InterPro; IPR011009; Kinase like.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.

```

DR Pfam; PF00069; Pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 2.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 686 AA; 77423 MW; 1B9F16E9FEFBAB82 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 686;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 944 SEKSSSSE 951  
|||  
Db 5 SEKSSSSE 12

Search completed: August 28, 2005, 11:08:01  
Job time : 191 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 10:49:27 ; Search time 170 Seconds  
(without alignments)  
2293.262 Million cell updates/sec

Title: US-10-078-531-2  
Perfect score: 1008  
Sequence: 1 MKKHLKTVALLTTTVSVVTH.....LGYTSVALLSLITAKKKY 1008

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 110 summaries

Database : A\_Geneseq\_16Dec04: \*  
1: Geneseqp1980s: \*  
2: Geneseqp1990s: \*  
3: Geneseqp2000s: \*  
4: Geneseqp2001s: \*  
5: Geneseqp2002s: \*  
6: Geneseqp2003as: \*  
7: Geneseqp2003bs: \*  
8: Geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1008	100.0	1008	5	ABP25813 Streptococ
2	1008	100.0	1008	5	ABP53347 Streptococ
3	1008	100.0	1008	8	ADR83923 S. pyogen
4	76	7.5	1055	4	AAU03612 Group B S
5	76	7.5	1055	5	ABP25812 Streptococ
6	76	7.5	1055	6	ABP56257 Serotype
7	21	2.1	21	5	ABP53348 Streptococ
8	9	0.9	106	4	AAW42172 Human pol
9	9	0.9	235	3	AGC43601 Arabidops
10	9	0.9	244	3	AGC43600 Arabidops
11	9	0.9	245	3	AGC43599 Arabidops
12	9	0.9	279	4	ABP92626 Human pro
13	9	0.9	279	7	ADJ69129 Human hea
14	9	0.9	340	3	AGC44240 Arabidops
15	9	0.9	349	3	AGC44239 Arabidops
16	9	0.9	361	4	AAW40386 Human pol
17	9	0.9	370	3	AGC44241 Arabidops
18	9	0.9	557	4	AAW48789 Human pro
19	9	0.9	557	5	ABP97502 Novel hum
20	9	0.9	581	3	AGC49069 Arabidops
21	9	0.9	583	8	ADJ50969 Human nov
22	9	0.9	590	3	AGC49068 Arabidops
23	9	0.9	590	5	ABP93671 Herbicida
24	9	0.9	608	7	ABM85378 Mouse pro
25	9	0.9	611	3	AGC49067 Arabidops

26	9	0.9	692	2	AAW93890	Human HG3
27	9	0.9	883	7	ADB80464	Ovarian c
28	9	0.9	883	7	ADN40012	Cancer/an
29	9	0.9	883	7	ADN39166	Cancer/an
30	9	0.9	907	2	AAW93889	Human HG3
31	9	0.9	907	3	AAW90682	Human G p
32	9	0.9	907	3	AAW90687	Human mut
33	9	0.9	907	6	ABP81968	Human G p
34	9	0.9	907	6	ABO06467	Human G-P
35	9	0.9	907	7	ADC22797	Human G p
36	9	0.9	907	7	ADC22783	Human G p
37	9	0.9	907	7	ADE59150	Human Pro
38	9	0.9	907	7	ADE59153	Human Pro
39	9	0.9	907	7	ADG42629	Human G p
40	9	0.9	907	7	ADG42628	Human G p
41	9	0.9	907	7	ADH14270	Mutated h
42	9	0.9	907	7	ADH14256	Human HG3
43	9	0.9	907	7	ADN40013	Cancer/an
44	9	0.9	907	7	ADN39531	Cancer/an
45	9	0.9	907	7	ADN39628	Cancer/an
46	9	0.9	907	8	ADO29408	Human GPC
47	9	0.9	907	8	ADQ80369	G protein
48	9	0.9	907	8	ADR67868	Human HG3
49	9	0.9	1145	7	ADF70480	Orphan re
50	8	0.8	147	4	ABG03518	Novel hum
51	8	0.8	196	7	ADG88441	Arabidops
52	8	0.8	240	7	ADM26567	Hyperther
53	8	0.8	264	7	ADG88453	Arabidops
54	8	0.8	268	5	ABH05713	Human cel
55	8	0.8	275	7	ADG88454	Arabidops
56	8	0.8	277	7	ADG88440	Arabidops
57	8	0.8	284	3	AA18240	Plasmodi
58	8	0.8	308	7	ADG88439	Arabidops
59	8	0.8	319	2	AAV03187	Rat Acid
60	8	0.8	350	2	AAV22200	S. frugip
61	8	0.8	356	6	ABU17197	Protein e
62	8	0.8	364	6	ADA33080	Acinetoba
63	8	0.8	411	5	ABP73447	Candida a
64	8	0.8	458	8	ADS28368	Bacterial
65	8	0.8	473	7	ADG88438	Arabidops
66	8	0.8	480	7	ABC08227	Rice prot
67	8	0.8	483	4	AAW92603	Human pro
68	8	0.8	488	8	ADJ48708	Oil-assoc
69	8	0.8	488	8	ADJ49363	Oil-assoc
70	8	0.8	488	8	ADJ48727	Oil-assoc
71	8	0.8	488	8	ADJ49362	Oil-assoc
72	8	0.8	496	7	ADC07838	Rice prot
73	8	0.8	500	8	ADJ50282	Oil-assoc
74	8	0.8	503	8	ADJ49407	Oil-assoc
75	8	0.8	503	8	ADJ48856	Oil-assoc
76	8	0.8	503	8	ADJ50184	Oil-assoc
77	8	0.8	519	4	AAW94834	Human pro
78	8	0.8	523	2	AAW67006	Recombina
79	8	0.8	526	2	AAW88327	Beta-amyl
80	8	0.8	527	8	ADJ48855	Oil-assoc
81	8	0.8	528	3	AGC31382	Arabidops
82	8	0.8	529	3	AGC31381	Arabidops
83	8	0.8	530	2	AAW67005	Barley be
84	8	0.8	531	2	AAW97613	Beta-amyl
85	8	0.8	533	8	ADJ49017	Oil-assoc
86	8	0.8	533	8	ADJ49016	Oil-assoc
87	8	0.8	535	2	AAW04261	Beta-amyl
88	8	0.8	535	2	AAW15746	Mutated b
89	8	0.8	535	8	ADJ49120	Oil-assoc
90	8	0.8	535	8	ADJ50393	Oil-assoc
91	8	0.8	535	8	ADJ50119	Oil-assoc
92	8	0.8	535	8	ADJ49037	Oil-assoc
93	8	0.8	535	8	ADJ50182	Oil-assoc
94	8	0.8	555	3	AAW07724	An Arabid
95	8	0.8	555	3	AGC31380	Arabidops
96	8	0.8	555	6	ABR40832	Arabidops
97	8	0.8	555	6	ABR40830	Arabidops
98	8	0.8	555	7	ADG25136	Thalecres



99 8 0.8 555 7 ADG88451 Arabidops  
100 8 0.8 555 7 ADG88431 Arabidops  
101 8 0.8 555 7 ADG88450 Arabidops  
102 8 0.8 555 8 ADO61539 Transcrip  
103 8 0.8 649 5 ABP69424  
104 8 0.8 660 4 ABG30817  
105 8 0.8 660 5 ABG38772 Human pep  
106 8 0.8 681 7 ADE59481 Rat Prote  
107 8 0.8 686 5 AAU80373 Human cel  
108 8 0.8 686 6 AAE34823 Protein #  
109 8 0.8 686 8 ADO19347 Human PRO  
110 8 0.8 688 2 AAY27053 Human pro

## ALIGNMENTS

RESULT 1  
ADP25813  
ID ABP25813 standard; protein; 1008 AA.  
XX  
AC ABP25813;  
XX  
DT 02-JUL-2002 (first entry)  
XX  
DE Streptococcus polypeptide SEQ ID NO 802.  
XX  
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX  
OS Streptococcus pyogenes.  
XX  
XX WO200234771-A2.  
PN  
XX  
PD 02-MAY-2002.  
XX  
PF 29-OCT-2001; 2001WO-GB004789.  
XX  
PR 27-OCT-2000; 2000GB-00026333.  
PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;  
PI Tettelin H;  
XX  
DR WPI: 2002-352536/38.  
DR N-PSDB; ABN6444.  
XX

PT New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
XX  
XX  
PS Claim 1; Page 3231; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying  
XX Streptococcus proteins  
SQ Sequence 1008 AA;  
Query Match 100.0%; Score 1008; DB 5; Length 1008;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MKGHLKTVLTLTVSVVTHNQEVFSLVKBPILKQTQASSISISGADVAESGSKLKLINE 60  
Db 1 MKGHLKTVLTLTVSVVTHNQEVFSLVKBPILKQTQASSISISGADVAESGSKLKLINE 60  
Qy 61 TSGPVDVDTVTLPSDKRTTPEKIKDNLAKGPREQLKAVTENTSEKQITSGSOLEOSKE 120  
Db 61 TSGPVDVDTVTLPSDKRTTPEKIKDNLAKGPREQLKAVTENTSEKQITSGSOLEOSKE 120  
Qy 121 SLSLNTKTVPSTSNWEICDFITKGNLTVLGLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180  
Db 121 SLSLNTKTVPSTSNWEICDFITKGNLTVLGLSKSGVEKLSQTDHLVLPQAADGTQLIQVAS 180  
Qy 181 FAPTDDKKTAAIAYTSRAGENGESIQLDVQDKEIINEGEVFNLSLLKKVTIPTGYKHIGQ 240  
Db 181 FAPTDDKKTAAIAYTSRAGENGESIQLDVQDKEIINEGEVFNLSLLKKVTIPTGYKHIGQ 240  
Qy 241 DAFVNDKNIAREVNLPESETISDYAFALALKOITDLPDNLKAIAGELAFPDNQITGKLSLP 300  
Db 241 DAFVNDKNIAREVNLPESETISDYAFALALKOITDLPDNLKAIAGELAFPDNQITGKLSLP 300  
Qy 301 RQLMRLAERAFKSNHIKTIEFRGNSLVKIGEASQDNDLSQMLPDGLKEKIESAFTGNP 360  
Db 301 RQLMRLAERAFKSNHIKTIEFRGNSLVKIGEASQDNDLSQMLPDGLKEKIESAFTGNP 360  
Qy 361 GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKMLEEDFTYQKNSVT 420  
Db 361 GDDHNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKMLEEDFTYQKNSVT 420  
Qy 421 GFSNKGLOKVRKNKLEIPKOHNGVTITEIGDNAPRVNDFONKTLRKLYDLEEVKLPSITR 480  
Db 421 GFSNKGLOKVRKNKLEIPKOHNGVTITEIGDNAPRVNDFONKTLRKLYDLEEVKLPSITR 480  
Qy 481 KIGAFAPQSNLKSFEASDDLEEKEGAFMNNRIETLELKDVLVTIGDAAFHINHIYAIV 540  
Db 481 KIGAFAPQSNLKSFEASDDLEEKEGAFMNNRIETLELKDVLVTIGDAAFHINHIYAIV 540  
Qy 541 LPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDLSEKQLTETIPVQAF 600  
Db 541 LPESVQIEGRSAFRQNGANNLI FMGSKVKTIGEMAFLSNRLEHLDLSEKQLTETIPVQAF 600  
Qy 601 SDNALKEVLLPASLKTIREERAFKKNHLKQLEVASALSIAFNALDDNDGDFONKVVVK 660  
Db 601 SDNALKEVLLPASLKTIREERAFKKNHLKQLEVASALSIAFNALDDNDGDFONKVVVK 660  
Qy 661 THNSYALADGEHFIVDPDKLSSITVDLEKILKIEGLDYSTLRQTQTQFRDMMTAGKA 720  
Db 661 THNSYALADGEHFIVDPDKLSSITVDLEKILKIEGLDYSTLRQTQTQFRDMMTAGKA 720  
Qy 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780  
Db 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780  
Qy 781 AYNNSAIKKNVKELEKELDLITGLVSGKGPLAAQATWVGYYLAKTLPPLPEYYIGLVNY 840  
Db 781 AYNNSAIKKNVKELEKELDLITGLVSGKGPLAAQATWVGYYLAKTLPPLPEYYIGLVNY 840  
Qy 841 FDKSGKLIYALDMDSTTIGEGQDAYGNPILNVDDNEGYPHALAVATLADYEGLDIKTILN 900  
Db 841 FDKSGKLIYALDMDSTTIGEGQDAYGNPILNVDDNEGYPHALAVATLADYEGLDIKTILN 900  
Qy 901 SKLSQLTISIRQVPTAAVHRAGIFQAIQNAABAEQQLLPKPGTHSEKSSSSSSANSKDRGL 960  
Db 901 SKLSQLTISIRQVPTAAVHRAGIFQAIQNAABAEQQLLPKPGTHSEKSSSSSSANSKDRGL 960

QY 961 QSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVALLSLITAIKKKKY 1008  
|||||  
Db 961 QSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVALLSLITAIKKKKY 1008  
|||||

## RESULT 2

ABP53347  
ID ABP53347 standard; protein; 1008 AA.

XX  
AC ABP53347;

XX 18-NOV-2002 (first entry)

XX

DE Streptococcus pyogenes BVH-P7 protein SEQ ID NO:2.

XX Streptococcus pyogenes; BVH-P7; antigen; group A Streptococcus; GAS;  
KW bacterial pathogen; vaccine; antibacterial; gene therapy; pharyngitis;  
KW streptococcal infection; erysipelas; impetigo; scarlet fever; infection;  
KW invasive disease; bacteraemia; necrotizing fasciitis; toxic shock.

XX Streptococcus pyogenes.

OS Key Location/Qualifiers

FH Peptide 1..21

FT Protein /label= signal

FT Protein 22..1008

FT Protein /label= mature\_BVH-P7

XX WO200266650-A2.

XX

PD 29-AUG-2002.

XX

PF 21-FEB-2002; 2002WO-CA000207.

XX

PR 21-FEB-2001; 2001US-0269840P.

XX

PA (SHIR-) SHIRE BIOCHEM INC.

XX

PI Martin D, Rioux S, Brodeur BR, Hamel J, Rheault P;

XX WPI: 2002-674948/72.

DR N-PSDB; ABQ81821.

XX

PT New polypeptide useful as a vaccine component for preventing, treating or  
PT diagnosing Streptococcus pyogenes infections, e.g. pharyngitis,  
PT erysipelas, impetigo, scarlet fever, bacteraemia, necrotizing fasciitis or  
PT toxic shock.

XX Claim 17; Fig 2; 52pp; English.

PS The present sequence represents Streptococcus pyogenes BVH-P7 protein  
CC (I). (I) has antibacterial activity and can be used in vaccines and gene  
CC therapy. The Streptococcus pyogenes BVH-P7 polypeptide is useful as a  
CC vaccine component for preventing, treating and/or diagnosing  
CC streptococcal infections, such as pharyngitis, erysipelas, impetigo,  
CC scarlet fever, and invasive diseases such as bacteraemia and necrotizing  
CC fasciitis, or toxic shock. A composition comprising the BVH-P7  
CC polypeptide is useful in the manufacture of a medicament for the  
CC prophylactic or therapeutic treatment of streptococcal infection. The BVH  
CC -P7 polynucleotide may be used in designing probes for the detection of  
CC Streptococcus in biological samples. The BVH-P7 polypeptide may also be  
CC used as immunogens for the production of antibodies against streptococcal  
CC infections

XX Sequence 1008 AA;

Query Match 100.0%; Score 1008; DB 5; Length 1008;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKHLKTVALTTTSSVTHNQEVSLVKEPILKQTQASSISGADYAESGSKLKINE 60  
|||||

Db 1 MKKHLKTVALTTTSSVTHNQEVSLVKEPILKQTQASSISGADYAESGSKLKINE 60  
|||||

QY 61 TSGPVDVTVDLFSKRTTPEKIKDNLAGPREQELKAVTENTESEKQITSGSQLEQSK 120  
|||||  
Db 61 TSGPVDVTVDLFSKRTTPEKIKDNLAGPREQELKAVTENTESEKQITSGSQLEQSK 120  
|||||

QY 121 SLSLNKTVPSSTSNWEICDFTKGNLTVLGLSKSGVEKLSQTDHLVLPQAADGTQIIOVAS 180  
|||||  
Db 121 SLSLNKTVPSSTSNWEICDFTKGNLTVLGLSKSGVEKLSQTDHLVLPQAADGTQIIOVAS 180  
|||||

QY 181 FAPTPDKKTAJAETSRAGENGEISQLDVDGKEIINEGEVFNYSLLKKVTIPTGYKHGQ 240  
|||||  
Db 181 FAPTPDKKTAJAETSRAGENGEISQLDVDGKEIINEGEVFNYSLLKKVTIPTGYKHGQ 240  
|||||

QY 241 DAFVNDKNIAEYNLPESLETISDYAFAPHAHLAKQIDLPDNLKAIGLAFDNOITGKLSLP 300  
|||||  
Db 241 DAFVNDKNIAEYNLPESLETISDYAFAPHAHLAKQIDLPDNLKAIGLAFDNOITGKLSLP 300  
|||||

QY 301 RQLMLABRAPKSNHIKTIETFRGNSLKVIGEASFQDNDLSQLMLPDGLEKIESEFTGNP 360  
|||||  
Db 301 RQLMLABRAPKSNHIKTIETFRGNSLKVIGEASFQDNDLSQLMLPDGLEKIESEFTGNP 360  
|||||

QY 361 GDDHYNRVLVLTWTKSGKNPSGLATENTYVNDPKSLWQESPEIDYTKWLEEDFTYKNSVT 420  
|||||  
Db 361 GDDHYNRVLVLTWTKSGKNPSGLATENTYVNDPKSLWQESPEIDYTKWLEEDFTYKNSVT 420  
|||||

QY 421 GFSNKGLOKVKRNKNLEIPKQHGVTITEIGDNAPRNVDVFNKTLRKYDLSEVKLPSTIR 480  
|||||  
Db 421 GFSNKGLOKVKRNKNLEIPKQHGVTITEIGDNAPRNVDVFNKTLRKYDLSEVKLPSTIR 480  
|||||

QY 481 KIGAFAPQSNNLKSPASDDLEEIKEGAFMNNRIETLELKKDKLVLTIGDAAPHINHIYAIV 540  
|||||  
Db 481 KIGAFAPQSNNLKSPASDDLEEIKEGAFMNNRIETLELKKDKLVLTIGDAAPHINHIYAIV 540  
|||||

QY 541 LPESVQEIGRSAFRONGANNLIEMGSKVKTIGEMAFLSNRLEHLDLSQKQITEIPVQAF 600  
|||||  
Db 541 LPESVQEIGRSAFRONGANNLIEMGSKVKTIGEMAFLSNRLEHLDLSQKQITEIPVQAF 600  
|||||

QY 601 SDNALKEVLLPASLKTIREEAFKKNHLKQEVASALSALSHIAFNALDDNDGDEQFNKVVVK 660  
|||||  
Db 601 SDNALKEVLLPASLKTIREEAFKKNHLKQEVASALSALSHIAFNALDDNDGDEQFNKVVVK 660  
|||||

QY 661 THNSYALADGEHFIVDPDKLSSTIVDLEKILKLEGLDYSITLRTTOTQFRDMTTAGKA 720  
|||||  
Db 661 THNSYALADGEHFIVDPDKLSSTIVDLEKILKLEGLDYSITLRTTOTQFRDMTTAGKA 720  
|||||

QY 721 LLSKSNLROGKQKFLQEAQFPLGRVLDLKAIAKAEKALVTCKATNGCOLLERSINKAVL 780  
|||||  
Db 721 LLSKSNLROGKQKFLQEAQFPLGRVLDLKAIAKAEKALVTCKATNGCOLLERSINKAVL 780  
|||||

QY 781 AYNNSAIKKANVXRLEKELDLLTGLVEGKGLAQATWVGYYLLKTPPLPEYYIGLVNY 840  
|||||  
Db 781 AYNNSAIKKANVXRLEKELDLLTGLVEGKGLAQATWVGYYLLKTPPLPEYYIGLVNY 840  
|||||

QY 841 FDKSGKLIYALDMSDTIGEGKDAYGNPILNVDENEGYPHALAVATLADYEGLDIKTILN 900  
|||||  
Db 841 FDKSGKLIYALDMSDTIGEGKDAYGNPILNVDENEGYPHALAVATLADYEGLDIKTILN 900  
|||||

QY 901 SKLSQLTSTIRQVPTAAVHRAGIQAIONAAAEQQLPKPGTHSEKSSSESANSKDRCL 960  
|||||  
Db 901 SKLSQLTSTIRQVPTAAVHRAGIQAIONAAAEQQLPKPGTHSEKSSSESANSKDRCL 960  
|||||

QY 961 QSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVALLSLITAIKKKKY 1008  
|||||  
Db 961 QSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVALLSLITAIKKKKY 1008  
|||||

RESULT 3

ADR83923

ID ADR83923 standard; protein; 1008 AA.

XX ADR83923;

XX 02-DEC-2004 (first entry)

XX S. pyogenes hyperimmune system reactive antigen Spy0843.  
 XX hyperimmune serum reactive antigen; vaccine; anticaline.  
 XX Streptococcus pyogenes.  
 XX WO2004078907-A2.  
 XX 16-SEP-2004.  
 XX 02-MAR-2004; 2004WO-EP002087.  
 XX 04-MAR-2003; 2003EP-00450061.  
 XX (INTE-) INTERCELL AG.  
 XX Meinke A, Nagy E, Winkler B, Gelbmann D;  
 XX WPI: 2004-653698/63.  
 XX N-PSDB; ADR83773.  
 XX New isolated nucleic acid molecules encoding hyperimmune serum-reactive  
 PT antigens from Streptococcus pyogenes, useful for diagnosing, preventing  
 PT and treating S. pyogenes infections.  
 XX Claim 14; SEQ ID NO 191; 145pp; English.  
 XX This invention describes a novel nucleic acid molecule encoding a  
 CC hyperimmune serum reactive antigen or its fragment from Streptococcus  
 CC pyogenes. The nucleic acid molecule or hyperimmune serum-reactive antigen  
 CC or its fragment are useful for the manufacture of a pharmaceutical  
 CC preparation, especially a vaccine, against S. pyogenes infection. In  
 CC addition, the hyperimmune serum reactive antigen or fragment is used for  
 CC the isolation and/or purification and/or identification of an interaction  
 CC partner of the hyperimmune serum reactive antigen or its fragment, for  
 CC the generation of a peptide (e.g. anticalines) binding to the antigen or  
 CC fragment, or for the manufacture of a functional nucleic acid selected  
 CC from aptamers and spiegelmers. The nucleic acid molecule may also be used  
 CC for the manufacture of functional ribonucleic acids, such as ribozymes,  
 CC antisense nucleic acids and siRNA. ADR83773-ADR84189 represent S.  
 CC pyogenes hyperimmune serum reactive antigens, fragments and the encoding  
 CC polynucleotide described in the invention.  
 XX Sequence 1008 AA;

Query Match 100.0%; Score 1008; DB 8; Length 1008;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKELKTVALTUTTVSVTHNQEVPSLVKPEILKQTOASSISGADYAESGSKLKINE 60  
 DB 1 MKKELKTVALTUTTVSVTHNQEVPSLVKPEILKQTOASSISGADYAESGSKLKINE 60

QY 61 TSGPVDVDTVDLFSDKRTTPEIKONLAKPREQLKAVTENTESEKQITSGSQLEQSK 120  
 DB 61 TSGPVDVDTVDLFSDKRTTPEIKONLAKPREQLKAVTENTESEKQITSGSQLEQSK 120

QY 121 SLSLNTKVPSTNSWEICDPIFKGNTLVGLSKGVEKLSQTDHLVLPSSQAADGTQLIQVAS 180  
 DB 121 SLSLNTKVPSTNSWEICDPIFKGNTLVGLSKGVEKLSQTDHLVLPSSQAADGTQLIQVAS 180

QY 181 FAFTPDKKTATAEYTSRAGENGEISQLDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240  
 DB 181 FAFTPDKKTATAEYTSRAGENGEISQLDVGKEIINEGEVFNYSLLKKVTIPTGYKHIGQ 240

QY 241 DAFVNKNIAEYNLPESLETISDYAFHALAKQIDLPNLKAIGELAFDNOITGKLSIP 300  
 DB 241 DAFVNKNIAEYNLPESLETISDYAFHALAKQIDLPNLKAIGELAFDNOITGKLSIP 300

QY 301 RQMLRAERAPKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360  
 DB 301 RQMLRAERAPKSNHIKTIIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEFTGNP 360

QY 361 GDDHNNRVVLWTSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420  
 DB 361 GDDHNNRVVLWTSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

QY 421 GFSNKGKQKVRKNKLEIPKQHNQVTITEIGDNAFRNVDFONKTLRKYDLEEVKLPSTIR 480  
 DB 421 GFSNKGKQKVRKNKLEIPKQHNQVTITEIGDNAFRNVDFONKTLRKYDLEEVKLPSTIR 480

QY 481 KIGAPAFQSNNLKSFEASDDLEIEKEGAFMNNRIETLELKDOKLVTIGDAAPHNHIAIV 540  
 DB 481 KIGAPAFQSNNLKSFEASDDLEIEKEGAFMNNRIETLELKDOKLVTIGDAAPHNHIAIV 540

QY 541 LPESVQEIGRSAFRONGANNLI PMGSKVKTIGENAFI SNRLEHLDSLSEKQLTPIPVQAF 600  
 DB 541 LPESVQEIGRSAFRONGANNLI PMGSKVKTIGENAFI SNRLEHLDSLSEKQLTPIPVQAF 600

QY 601 SDNALKEVLLPASLKTITREERAFKKNHLKQLEVASALSHIAFNALDDNDGDQFONKVVVK 660  
 DB 601 SDNALKEVLLPASLKTITREERAFKKNHLKQLEVASALSHIAFNALDDNDGDQFONKVVVK 660

QY 661 THNSYALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLRQTQTQPRDMTTAGKA 720  
 DB 661 THNSYALADGEHFIVDPDKLSSTIVDLKILKLEGLDYSTLRQTQTQPRDMTTAGKA 720

QY 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780  
 DB 721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIKAEKALVTKKATKNGQLLERSINKAVL 780

QY 781 AYNNSAIKKNVKKLEKELDLTLGLVSGKGLAQATWVGYYLLKTPPLPEYIYGLNVY 840  
 DB 781 AYNNSAIKKNVKKLEKELDLTLGLVSGKGLAQATWVGYYLLKTPPLPEYIYGLNVY 840

QY 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDSNDEGYHALAVATLADYEGLDIKTILN 900  
 DB 841 FDKSGKLIYALDMSDTTIGEGQKDAYGNPILNVDSNDEGYHALAVATLADYEGLDIKTILN 900

QY 901 SKLSQLTSIROVPTAAVHRAGIFQAIQNAABAEQLLPKPTHSEKSSSSSANSKDRGL 960  
 DB 901 SKLSQLTSIROVPTAAVHRAGIFQAIQNAABAEQLLPKPTHSEKSSSSSANSKDRGL 960

QY 961 QSNPKTNRGRHSAILPRTGSKGSFVYGLYTSVALLSLITAIKKKKY 1008  
 DB 961 QSNPKTNRGRHSAILPRTGSKGSFVYGLYTSVALLSLITAIKKKKY 1008

RESULT 4  
 AAU03612  
 ID AAU03612 standard; protein; 1055 AA.  
 XX  
 AC AAU03612;  
 XX  
 DT 12-SEP-2001 (first entry)  
 XX  
 DE Group B Streptococcus antigenic protein, ID-87.  
 XX  
 KW Group B Streptococcus; encapsulated bacterium; therapeutic; sepsis;  
 KW meningitis; neonate; antigenic; vaccine; infection; genital tract;  
 KW capsid polysaccharide vaccination.  
 XX  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200132882-A2.  
 XX  
 PD 10-MAY-2001.  
 XX  
 PF 07-SEP-2000; 2000WO-GB003437.  
 XX  
 PR 07-SEP-1999; 99GB-00021125.  
 XX  
 PA (MICR-) MICROBIAL TECHNIQS LTD.  
 XX  
 PI Le Page RWP, Wells JM, Hanniffy SB;

XX WPI; 2001-316444/33.  
 DR N-PSDB; AAS07029.  
 XX  
 PT New polypeptides derived from Streptococcus agalactiae are useful to  
 PT provide detection of, and vaccination against, Group B Streptococcus  
 PT infections, particularly to prevent infection in neonates.  
 XX  
 PS Claim 1; Fig 1; 178pp; English.  
 XX  
 CC AAU03601-AAU03722 represent Group B Streptococcus (Streptococcus  
 CC agalactiae) amino acid sequences of the invention. S. agalactiae is an  
 CC encapsulated bacterium which is a major pathogen of humans causing sepsis  
 CC and meningitis in neonates as well as adults. The S. agalactiae antigenic  
 CC polypeptides are used to vaccinate against Group B Streptococcus  
 CC infections, particularly to prevent infection in new born children  
 CC arising from the maternal genital tract. An immunogenic composition is  
 CC useful in the preparation of a medicament for the treatment or  
 CC prophylaxis of Group B Streptococcus infection. The invention does not  
 CC have the disadvantages of varied response rate associated with prior art  
 CC capsid polysaccharide vaccination against Group B Streptococcus  
 XX  
 SQ Sequence 1055 AA;  
 Query Match 7.5%; Score 76; DB 4; Length 1055;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-65;  
 Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 805 LVVEGKPLAQATWVGQVLLKTPPLPEYIYGLNVYFDKSGKLIYALDMSDTIGGQKDA 864  
 DB 806 LVVEGKPLAQATWVGQVLLKTPPLPEYIYGLNVYFDKSGKLIYALDMSDTIGGQKDA 865  
 QY 865 YGNPILNVDEDNDEGYH 880  
 DB 866 YGNPILNVDEDNDEGYH 881  
 RESULT 5  
 ABP25812  
 ID ABP25812 standard; protein; 1055 AA.  
 XX  
 AC ABP25812;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polypeptide SEQ ID NO 800.  
 XX  
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-GB0004789.  
 XX  
 PR 27-OCT-2000; 2000GB-00026333.  
 PR 24-NOV-2000; 2000GB-00028727.  
 PR 07-MAR-2001; 2001GB-00005640.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Telford J, Massignani V, Margarit Y Rosi, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 DR WPI; 2002-352536/38.  
 DR N-PSDB; ABN66443.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection or

PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 PT detecting a compound that binds to the protein.  
 XX  
 PS Claim 1; Page 3230-3231; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins  
 XX  
 SQ Sequence 1055 AA;  
 Query Match 7.5%; Score 76; DB 5; Length 1055;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-65;  
 Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 805 LVVEGKPLAQATWVGQVLLKTPPLPEYIYGLNVYFDKSGKLIYALDMSDTIGGQKDA 864  
 DB 806 LVVEGKPLAQATWVGQVLLKTPPLPEYIYGLNVYFDKSGKLIYALDMSDTIGGQKDA 865  
 QY 865 YGNPILNVDEDNDEGYH 880  
 DB 866 YGNPILNVDEDNDEGYH 881  
 RESULT 6  
 ABP56257  
 ID ABP56257 standard; protein; 1055 AA.  
 XX  
 AC ABP56257;  
 XX  
 DT 28-MAR-2003 (first entry)  
 XX  
 DE Serotype III group B Streptococcus strain COH1 BVH-A4 SEQ ID NO:2.  
 XX  
 KW Serotype III group B streptococcus strain COH1; BVH-A4; streptococcus;  
 KW antibiotic; immunostimulant; vaccine; bacterial infection; sepsis;  
 KW meningitis; pneumonia; cellulitis; osteomyelitis; septic arthritis;  
 KW endocarditis; epiglottitis; osteomyelitis; amniotitis; endometritis;  
 KW cellulitis; fasciitis; bacteraemia; urosepsis; peritonitis; emphysema;  
 KW mastitis; streptococcal infection.  
 XX  
 OS Streptococcus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..22  
 FT Protein /label= signal  
 FT 23..1055  
 FT /label= BVH-A4  
 XX  
 PN WO200288178-A2.  
 XX  
 PD 07-NOV-2002.  
 XX  
 PF 02-MAY-2002; 2002WO-CA000664.  
 XX  
 PR 02-MAY-2001; 2001US-0287712P.  
 XX  
 PA (SHIR-) SHIRE BIOCHEM INC.  
 XX

PI Martin D, Hamel J, Brodeur BR, Rioux S, Boyer M;  
 XX WPI; 2003-120461/11.  
 DR N-PSDB; ABZ21973.  
 XX  
 PT New BVH-A4 proteins and genes from serotype III Group B streptococcus,  
 XX useful for treating or preventing streptococcal infection in infants,  
 PT pregnant women, non-pregnant adults (e.g. pneumonia), or members of dairy  
 PT herd (mastitis).  
 XX  
 PS Claim 18; Fig 2; 60pp; English.  
 XX  
 CC The present sequence represents a BVH-A4 protein from serotype III Group  
 CC B streptococcus (GBS) strain COH1 (designated GBS-BVH-A4) (I). (I) has  
 CC antibiotic and immunostimulant activities, and can be used in polypeptide  
 CC therapy and in vaccine production. (I) can be used for the therapeutic or  
 CC prophylactic treatment of GBS bacterial infection in a host susceptible  
 CC to GBS infection. In particular, (I) is useful for treating or preventing  
 CC GBS infection in a neonate or infant (e.g. sepsis, meningitis, pneumonia,  
 CC cellulitis, osteomyelitis, septic arthritis, endocarditis or  
 CC epiglottitis), in a pregnant woman (e.g. mild urinary tract infection to  
 CC life-threatening sepsis and meningitis, osteomyelitis, endocarditis,  
 CC amniotitis, endometritis, wound infection (post-caesarean or post-  
 CC episiotomy), cellulitis or fasciitis), in a non-pregnant adult (e.g.  
 CC bacteraemia, skin or soft tissue infection, pneumonia, urosepsis,  
 CC endocarditis, peritonitis, meningitis or emphysema), or in a member of  
 CC dairy herd (e.g. mastitis). A composition comprising (I) or (I) can also  
 CC be useful for treating or preventing streptococcal infection  
 XX  
 SQ Sequence 1055 AA;  
 Query Match 7.5%; Score 76; DB 6; Length 1055;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-65;  
 Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 805 LVEGKGLAQAATMGVGVLLKTPLEPYIGLVNVPFDSKGLIYALDMSDTIGSGQKDA 864  
 DB 806 LVEGKGLAQAATMGVGVLLKTPLEPYIGLVNVPFDSKGLIYALDMSDTIGSGQKDA 865  
 QY 865 YGNPILNVDENEGYH 880  
 DB 866 YGNPILNVDENEGYH 881  
 RESULT 7  
 ABP53348  
 ID ABP53348 standard; peptide; 21 AA.  
 AC ABP53348;  
 XX  
 DT 18-NOV-2002 (first entry)  
 XX  
 DE Streptococcus pyogenes BVH-P7 signal peptide.  
 XX  
 KW Streptococcus pyogenes; BVH-P7; antigen; group A Streptococcus; GAS;  
 KW bacterial pathogen; vaccine; antibacterial; gene therapy; pharyngitis;  
 KW streptococcal infection; erysipelas; impetigo; scarlet fever; infection;  
 KW invasive disease; bacteraemia; necrotizing fasciitis; toxic shock.  
 XX  
 OS Streptococcus pyogenes.  
 XX  
 PN WO20026650-A2.  
 XX  
 PD 29-AUG-2002.  
 XX  
 PF 21-FEB-2002; 2002WO-CA000207.  
 XX  
 PR 21-FEB-2001; 2001US-0269840P.  
 XX  
 PA (SHIR-) SHIRE BIOCHEM INC.  
 XX  
 PI Martin D, Rioux S, Brodeur BR, Hamel J, Rheault P;  
 XX

DR WPI; 2002-674948/72.  
 XX  
 PT New polypeptide useful as a vaccine component for preventing, treating or  
 PT diagnosing Streptococcus pyogenes infections, e.g. pharyngitis,  
 PT erysipelas, impetigo, scarlet fever, bacteraemia, necrotizing fasciitis or  
 XX toxic shock.  
 PS Example 1; Page 25; 52pp; English.  
 XX  
 CC The present invention describes Streptococcus pyogenes BVH-P7 protein  
 CC (I). (I) has antibacterial activity and can be used in vaccines and gene  
 CC therapy. The Streptococcus pyogenes BVH-P7 polypeptide is useful as a  
 CC vaccine component for preventing, treating and/or diagnosing  
 CC streptococcal infections, such as pharyngitis, erysipelas, impetigo,  
 CC scarlet fever, and invasive diseases such as bacteraemia and necrotizing  
 CC fasciitis, or toxic shock. A composition comprising the BVH-P7  
 CC polypeptide is useful in the manufacture of a medicament for the  
 CC prophylactic or therapeutic treatment of streptococcal infection. The BVH  
 CC -P7 polynucleotide may be used in designing probes for the detection of  
 CC Streptococcus in biological samples. The BVH-P7 polypeptide may also be  
 CC used as immunogens for the production of antibodies against streptococcal  
 CC infections. The present sequence represents the signal peptide of S.  
 CC pyogenes BVH-P7, which is given in an example from the present invention  
 XX  
 SQ Sequence 21 AA;  
 Query Match 2.1%; Score 21; DB 5; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-12;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKGHLKTVALTTLTVSVVTHN 21  
 DB 1 MKGHLKTVALTTLTVSVVTHN 21  
 RESULT 8  
 AAM42172  
 ID AAM42172 standard; protein; 106 AA.  
 XX  
 AC AAM42172;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 7103.  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US034263.  
 XX  
 PR 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 20-JUN-2000; 2000US-00598042.  
 PR 19-JUL-2000; 2000US-00620312.  
 PR 03-AUG-2000; 2000US-00653450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 XX

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Drmanac RT;  
 XX WPI: 2001-442253/47.  
 DR N-PSDB; AA161328.  
 XX Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX Example 2; SEQ ID NO 7103; 10078pp; English.  
 XX The invention relates to human nucleic acids (AA157798-AA161369) and the  
 CC encoded polypeptides (AA38642-AA42213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification  
 XX Sequence 106 AA;  
 SQ

Query Match 0.9%; Score 9; DB 4; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 6.3;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSES 952  
 Db |||||  
 36 SEKSSSES 44

RESULT 9  
 AAG43601  
 ID AAG43601 standard; protein; 235 AA.  
 XX AC AAG43601;  
 XX 18-OCT-2000 (first entry)  
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 54515.  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX Arabidopsis thaliana.  
 XX EP1033405-A2.  
 XX 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-00301439.  
 XX 25-FEB-1999; 99US-0121825P.  
 PR 05-MAR-1999; 99US-0123180P.  
 PR 09-MAR-1999; 99US-0123548P.  
 PR 23-MAR-1999; 99US-0125788P.  
 PR 25-MAR-1999; 99US-0126264P.  
 PR 29-MAR-1999; 99US-0126785P.  
 PR 01-APR-1999; 99US-0127462P.  
 PR 06-APR-1999; 99US-0128234P.  
 PR 08-APR-1999; 99US-0128714P.  
 PR 16-APR-1999; 99US-0128845P.  
 PR 19-APR-1999; 99US-0130077P.  
 PR 21-APR-1999; 99US-0130449P.  
 PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.  
 PR 28-APR-1999; 99US-0131449P.  
 PR 30-APR-1999; 99US-0132048P.  
 PR 04-APR-1999; 99US-0132407P.  
 PR 04-MAY-1999; 99US-0132484P.  
 PR 05-MAY-1999; 99US-0132485P.  
 PR 06-MAY-1999; 99US-0132486P.  
 PR 06-MAY-1999; 99US-0132487P.  
 PR 07-MAY-1999; 99US-0132863P.  
 PR 11-MAY-1999; 99US-0134256P.  
 PR 14-MAY-1999; 99US-0134218P.  
 PR 14-MAY-1999; 99US-0134219P.  
 PR 14-MAY-1999; 99US-0134221P.  
 PR 18-MAY-1999; 99US-0134370P.  
 PR 18-MAY-1999; 99US-0134768P.  
 PR 19-MAY-1999; 99US-0134941P.  
 PR 20-MAY-1999; 99US-0135124P.  
 PR 21-MAY-1999; 99US-0135353P.  
 PR 24-MAY-1999; 99US-0135629P.  
 PR 25-MAY-1999; 99US-0136021P.  
 PR 27-MAY-1999; 99US-0136392P.  
 PR 28-MAY-1999; 99US-0136782P.  
 PR 01-JUN-1999; 99US-0137222P.  
 PR 03-JUN-1999; 99US-0137528P.  
 PR 04-JUN-1999; 99US-0137502P.  
 PR 07-JUN-1999; 99US-0137724P.  
 PR 08-JUN-1999; 99US-0138094P.  
 PR 10-JUN-1999; 99US-0138540P.  
 PR 10-JUN-1999; 99US-0138847P.  
 PR 14-JUN-1999; 99US-0139119P.  
 PR 16-JUN-1999; 99US-0139452P.  
 PR 16-JUN-1999; 99US-0139453P.  
 PR 17-JUN-1999; 99US-0139492P.  
 PR 18-JUN-1999; 99US-0139454P.  
 PR 18-JUN-1999; 99US-0139455P.  
 PR 18-JUN-1999; 99US-0139456P.  
 PR 18-JUN-1999; 99US-0139457P.  
 PR 18-JUN-1999; 99US-0139458P.  
 PR 18-JUN-1999; 99US-0139459P.  
 PR 18-JUN-1999; 99US-0139460P.  
 PR 18-JUN-1999; 99US-0139461P.  
 PR 18-JUN-1999; 99US-0139462P.  
 PR 18-JUN-1999; 99US-0139463P.  
 PR 18-JUN-1999; 99US-0139750P.  
 PR 18-JUN-1999; 99US-0139763P.  
 PR 21-JUN-1999; 99US-0139817P.  
 PR 22-JUN-1999; 99US-0139899P.  
 PR 23-JUN-1999; 99US-0140353P.  
 PR 23-JUN-1999; 99US-0140354P.  
 PR 24-JUN-1999; 99US-0140695P.  
 PR 28-JUN-1999; 99US-0140823P.  
 PR 29-JUN-1999; 99US-0140991P.  
 PR 30-JUN-1999; 99US-0141287P.  
 PR 01-JUL-1999; 99US-0141842P.  
 PR 01-JUL-1999; 99US-0142154P.  
 PR 02-JUL-1999; 99US-0142055P.  
 PR 06-JUL-1999; 99US-0142390P.  
 PR 08-JUL-1999; 99US-0142803P.  
 PR 09-JUL-1999; 99US-0142920P.  
 PR 12-JUL-1999; 99US-0142977P.  
 PR 13-JUL-1999; 99US-0143542P.  
 PR 14-JUL-1999; 99US-0143624P.  
 PR 15-JUL-1999; 99US-0144005P.  
 PR 16-JUL-1999; 99US-0144085P.  
 PR 16-JUL-1999; 99US-0144086P.  
 PR 19-JUL-1999; 99US-0144325P.  
 PR 19-JUL-1999; 99US-0144331P.  
 PR 19-JUL-1999; 99US-0144332P.  
 PR 19-JUL-1999; 99US-0144333P.  
 PR 19-JUL-1999; 99US-0144334P.  
 PR 19-JUL-1999; 99US-0144335P.  
 PR 20-JUL-1999; 99US-0144352P.  
 PR 20-JUL-1999; 99US-0144632P.



```
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147418P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.

PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160788P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 0.9%; Score 9; DB 3; Length 235;
Best Local Similarity 100.0%; Pred.No.13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAREQLL 937
DB 123 AAAREQLL 131
|||||||

RESULT 10
AAG43600
ID AAG43600 standard; protein; 244 AA.
XX
AC AAG43600;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 54514.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
```



```
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 0.9%; Score 9; DB 3; Length 244;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAEAEQQL 937
DB 132 AAAEAEQQL 140

RESULT 11
AAG43599
ID AAG43599 standard; protein; 245 AA.
AC AAG43599;
XX
XX
XX 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 54513.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridization assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0134407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132662P.
PR 11-MAY-1999; 99US-0134258P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.

18-MAY-1999; 99US-0134768P.
19-MAY-1999; 99US-0134941P.
20-MAY-1999; 99US-0135124P.
21-MAY-1999; 99US-0135353P.
24-MAY-1999; 99US-0135629P.
25-MAY-1999; 99US-0136021P.
27-MAY-1999; 99US-0136392P.
28-MAY-1999; 99US-0136782P.
01-JUN-1999; 99US-0137222P.
03-JUN-1999; 99US-0137528P.
04-JUN-1999; 99US-0137502P.
07-JUN-1999; 99US-0137724P.
08-JUN-1999; 99US-0138094P.
10-JUN-1999; 99US-0138540P.
10-JUN-1999; 99US-0138847P.
14-JUN-1999; 99US-0139119P.
16-JUN-1999; 99US-0139452P.
16-JUN-1999; 99US-0139453P.
17-JUN-1999; 99US-0139492P.
18-JUN-1999; 99US-0139454P.
18-JUN-1999; 99US-0139455P.
18-JUN-1999; 99US-0139456P.
18-JUN-1999; 99US-0139457P.
18-JUN-1999; 99US-0139458P.
18-JUN-1999; 99US-0139459P.
18-JUN-1999; 99US-0139460P.
18-JUN-1999; 99US-0139461P.
18-JUN-1999; 99US-0139462P.
18-JUN-1999; 99US-0139463P.
18-JUN-1999; 99US-0139750P.
18-JUN-1999; 99US-0139763P.
21-JUN-1999; 99US-0139817P.
22-JUN-1999; 99US-0139899P.
23-JUN-1999; 99US-0140353P.
23-JUN-1999; 99US-0140354P.
24-JUN-1999; 99US-0140695P.
28-JUN-1999; 99US-0140823P.
29-JUN-1999; 99US-0140931P.
30-JUN-1999; 99US-0141287P.
01-JUL-1999; 99US-0141842P.
01-JUL-1999; 99US-0142154P.
02-JUL-1999; 99US-0142055P.
06-JUL-1999; 99US-0142390P.
08-JUL-1999; 99US-0142803P.
09-JUL-1999; 99US-0142920P.
12-JUL-1999; 99US-0142977P.
13-JUL-1999; 99US-0143542P.
14-JUL-1999; 99US-0143624P.
15-JUL-1999; 99US-0144005P.
16-JUL-1999; 99US-0144085P.
16-JUL-1999; 99US-0144086P.
19-JUL-1999; 99US-0144325P.
19-JUL-1999; 99US-0144331P.
19-JUL-1999; 99US-0144332P.
19-JUL-1999; 99US-0144333P.
19-JUL-1999; 99US-0144334P.
19-JUL-1999; 99US-0144335P.
20-JUL-1999; 99US-0144352P.
20-JUL-1999; 99US-0144632P.
20-JUL-1999; 99US-0144884P.
21-JUL-1999; 99US-0144814P.
21-JUL-1999; 99US-0145086P.
21-JUL-1999; 99US-0145088P.
22-JUL-1999; 99US-0145085P.
22-JUL-1999; 99US-0145087P.
22-JUL-1999; 99US-0145089P.
22-JUL-1999; 99US-0145192P.
23-JUL-1999; 99US-0145145P.
23-JUL-1999; 99US-0145218P.
23-JUL-1999; 99US-0145224P.
26-JUL-1999; 99US-0145276P.
27-JUL-1999; 99US-0145913P.
27-JUL-1999; 99US-0145918P.
```

```
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159638P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.

PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      0.9%; Score 9; DB 3; Length 245;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAREAQLL 937
   |||||
Db 133 AAAREAQLL 141

RESULT 12
AAB92626
ID AAB92626 standard; protein; 279 AA.
AC AAB92626;
XX
XX 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:10923.
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX
XX EP1074617-A2.
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
XX Claim 8; SEQ ID NO 10923; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by
```

CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention

XX  
SQ Sequence 279 AA;

Query Match 0.9%; Score 9; DB 4; Length 279;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSES 952  
Db 18 SEKSSSES 26

RESULT 13  
ADJ69129  
ID ADJ69129 standard; protein; 279 AA.

XX AC ADJ69129;

XX DT 06-MAY-2004 (first entry)

XX DE Human heat mitochondrial protein as a therapeutic target SeqID935.

XX KW mitochondrial; human; screening assay; diabetes mellitus;  
XX KW Huntington's disease; osteoarthritis;  
XX KW Leber's hereditary optic neuropathy; LHON;  
XX KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
XX KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;  
XX KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;  
XX KW osteopathic; ophthalmological; cytostatic.

XX OS Homo sapiens.

XX PN WO2003087768-A2.

XX PD 23-OCT-2003.

XX PF 04-APR-2003; 2003WO-US010870.

XX PR 12-APR-2002; 2002US-0372843P.

XX PR 17-JUN-2002; 2002US-0389987P.

XX PR 20-SEP-2002; 2002US-0412418P.

XX PA (MITO-) MITOKOR.

XX PA (BUCK-) BUCK INST AGE RES.

XX PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;

XX PI Warnock DE;

XX DR WPI; 2003-845369/78.

XX PT Identifying a mitochondrial target for drug screening assays and for

XX PT treating diseases associated with altered mitochondrial function,

XX PT comprises detecting a modified polypeptide in a sample and correlating

XX PT with the disease.

XX PS Claim 1; SEQ ID NO 935; 180pp; English.

XX CC This invention relates to novel mitochondrial targets that can be used  
CC for therapeutic intervention in treating a disease associated with  
CC altered mitochondrial function. Specifically, it refers to a method for  
CC identifying proteins of the human heart mitochondrial proteome that are  
CC useful for drug screening assays, as well as therapeutic targets. The  
CC present invention describes a method for identifying such proteins that  
CC can be used in the treatment of various diseases associated with altered  
CC mitochondrial function including diabetes mellitus, Huntington's disease,  
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial  
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy

CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
CC compositions have neuroprotective, nootropic, antidiabetic,  
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
CC cytostatic activities. This polypeptide sequence is a human heart  
CC mitochondrial protein of the invention.

XX SQ Sequence 279 AA;

Query Match 0.9%; Score 9; DB 7; Length 279;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSES 952  
Db 18 SEKSSSES 26

RESULT 14

AAG44240

ID AAG44240 standard; protein; 340 AA.

XX AC AAG44240;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 55391.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

XX PR 05-MAR-1999; 99US-0123180P.

XX PR 09-MAR-1999; 99US-0123548P.

XX PR 23-MAR-1999; 99US-0125788P.

XX PR 25-MAR-1999; 99US-0126284P.

XX PR 29-MAR-1999; 99US-0126785P.

XX PR 01-APR-1999; 99US-0127462P.

XX PR 08-APR-1999; 99US-0128234P.

XX PR 16-APR-1999; 99US-0128714P.

XX PR 19-APR-1999; 99US-0129845P.

XX PR 21-APR-1999; 99US-0130077P.

XX PR 23-APR-1999; 99US-0130449P.

XX PR 28-APR-1999; 99US-0130891P.

XX PR 30-APR-1999; 99US-0131449P.

XX PR 30-APR-1999; 99US-0132048P.

XX PR 04-MAY-1999; 99US-0132407P.

XX PR 05-MAY-1999; 99US-0132484P.

XX PR 06-MAY-1999; 99US-0132485P.

XX PR 07-MAY-1999; 99US-0132486P.

XX PR 11-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132488P.

XX PR 14-MAY-1999; 99US-0134218P.

XX PR 14-MAY-1999; 99US-0134221P.

XX PR 14-MAY-1999; 99US-0134370P.

XX PR 18-MAY-1999; 99US-0134768P.

XX PR 19-MAY-1999; 99US-0134941P.

XX PR 20-MAY-1999; 99US-0135124P.

XX PR 21-MAY-1999; 99US-0135353P.

XX PR 24-MAY-1999; 99US-0135629P.

XX PR 25-MAY-1999; 99US-0136021P.

XX PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138840P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140595P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147418P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159291P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 22-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.



```
Query Match          0.9%; Score 9; DB 3; Length 340;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAARAEQLL 937
Db 123 AAARAEQLL 131

RESULT 15
AAG44239
ID AAG44239 standard; protein; 349 AA.
XX
AC AAG44239;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55390.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0128785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130443P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144633P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147322P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
```

PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151338P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.  
  
Query Match 0.9%; Score 9; DB 3; Length 349;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 929 AAAAEQQLL 937  
|||||||

Db 132 AAAAEQQLL 140  
  
RESULT 16  
AAW40386  
ID AAM40386 standard; protein; 361 AA.  
XX  
AC AAM40386;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 3531.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US034263.  
XX  
PR 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 20-JUN-2000; 2000US-00598042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;  
XX  
WPI; 2001-442253/47.  
DR N-PSDB; AAI59542.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
PS Example 6; SEQ ID NO 3531; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM38642-AAW42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity; chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
SQ Sequence 361 AA;  
  
Query Match 0.9%; Score 9; DB 4; Length 361;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY      944 SEKSSSES 952
Db      |||||
        18 SEKSSSES 26

RESULT 17
AAG44241
ID AAG44241 standard; protein; 370 AA.
XX
AC AAG44241;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55392.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 200DEP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0134048P.
PR 30-APR-1999; 99US-0134407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142330P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147935P.
PR 09-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
```

PR 13-AUG-1999; 99US-0148684P.  
 PR 16-AUG-1999; 99US-0149368P.  
 PR 17-AUG-1999; 99US-0149175P.  
 PR 18-AUG-1999; 99US-0149426P.  
 PR 20-AUG-1999; 99US-0149722P.  
 PR 20-AUG-1999; 99US-0149723P.  
 PR 20-AUG-1999; 99US-0149929P.  
 PR 23-AUG-1999; 99US-0149902P.  
 PR 23-AUG-1999; 99US-0149930P.  
 PR 25-AUG-1999; 99US-0150566P.  
 PR 26-AUG-1999; 99US-0150884P.  
 PR 27-AUG-1999; 99US-0151065P.  
 PR 27-AUG-1999; 99US-0151066P.  
 PR 30-AUG-1999; 99US-0151080P.  
 PR 31-AUG-1999; 99US-0151303P.  
 PR 31-AUG-1999; 99US-0151438P.  
 PR 01-SEP-1999; 99US-0151930P.  
 PR 07-SEP-1999; 99US-0152363P.  
 PR 10-SEP-1999; 99US-0153070P.  
 PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154039P.  
 PR 20-SEP-1999; 99US-0154779P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 23-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155659P.  
 PR 28-SEP-1999; 99US-0156458P.  
 PR 29-SEP-1999; 99US-0156596P.  
 PR 04-OCT-1999; 99US-0157117P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157865P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0158293P.  
 PR 13-OCT-1999; 99US-0158294P.  
 PR 13-OCT-1999; 99US-0158295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161992P.  
 PR 28-OCT-1999; 99US-0161993P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 0.9%; Score 9; DB 3; Length 370;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 AAAAEAEQLL 937  
 DB 153 AAAAEAEQLL 161

RESULT 18

AAB48789  
 ID AAB48789 standard; protein; 557 AA.  
 XX  
 AC AAB48789;  
 XX  
 DT 09-MAR-2001 (first entry)  
 XX  
 DE Human prostate cancer-predisposing protein, CA7 CG04.  
 XX  
 KW Human; CA7 CG04 gene; prostate cancer-predisposing gene;  
 KW chromosome 1 HPC1 region; inherited early onset prostate cancer;  
 KW guanosine exchange factor protein; GEF homologue; ras activator;  
 KW diagnosis; anticancer drug screening; mutation screening.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200069879-A2.  
 XX  
 PD 23-NOV-2000.  
 XX  
 PF 11-MAY-2000; 2000WO-US012917.  
 XX  
 PR 14-MAY-1999; 99US-0134209P.  
 XX  
 PA (MYRI-) MYRIAD GENETICS INC.  
 PA (HOSP-) HOSPITAL FOR SICK CHILDREN.  
 XX  
 PI Tavtigian SV, Swedlund B, Simard J, Rommens JM;  
 DR WPI; 2001-016208/02.  
 DR N-PSDB; AAC87385.  
 XX  
 PT Novel human prostate cancer predisposing gene useful for diagnosis and  
 PT prognosis of cancer, especially, prostate cancer and for screening drugs  
 PT for cancer therapy.  
 XX  
 PS Claim 1; Page 91-92; 96pp; English.  
 CC The invention relates to a human prostate cancer-predisposing gene,  
 CC designated CA7 CG04 (cDNA given in AAC87385), and the CA7 CG04 protein  
 CC (AAB48789). The CA7 CG04 protein has homology with guanosine exchange  
 CC factor (GEF) proteins, and is therefore thought to function as an  
 CC activator of ras. The CA7 CG04 gene is located on chromosome 1 in the  
 CC HPC1 region, which is linked with inherited early onset prostate cancer.  
 CC The invention also relates to exons 1-19 of the CA7 CG04 gene (AAC87386-  
 CC C87404), allelic variants and mutants of the CA7 CG04 gene, expression  
 CC vectors and host cells comprising a CA7 CG04 DNA, recombinant production  
 CC of the CA7 CG04 protein, anti-CA7 CG04 antibodies, CA7 CG04 primers and  
 CC probes, and methods of screening for potential anticancer drugs which can  
 CC inhibit the ability of a mutant CA7 CG04 to activate ras. The CA7 CG04  
 CC cDNA and protein are useful for screening potential anticancer drugs and  
 CC for screening the CA7 CG04 gene for mutations which are useful in  
 CC diagnosing a predisposition to cancer, especially prostate cancer. The  
 CC present sequence represents human CA7 CG04 protein  
 XX  
 SQ Sequence 557 AA;

Query Match 0.9%; Score 9; DB 4; Length 557;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSES 952  
 DB 18 SEKSSSES 26

RESULT 19  
 ABB97502  
 ID ABB97502 standard; protein; 557 AA.  
 XX  
 AC ABB97502;  
 XX

DT 27-JUN-2002 (first entry)

XX Novel human protein SEQ ID NO: 770.  
XX  
XX Human; antianemic; vulnerary; antiinflammatory; immunomodulator;  
KW antiferility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
KW neuroprotective; antiparkinsonian; protein therapy; EST;  
KW expressed sequence tag.  
XX  
OS Homo sapiens.  
XX  
XX WO200222660-A2.  
XX  
XX PD  
XX 21-MAR-2002.  
XX  
PF 10-SEP-2001; 2001WO-0026015.  
XX  
XX 11-SEP-2000; 2000US-00659671.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX  
XX WPI; 2002-292408/33.  
DR N-PSDB; ABN32688.  
XX  
XX An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis.  
XX  
XX Claim 20; SEQ ID NO 770; 509pp; English.  
XX  
XX The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke  
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.  
CC rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a protein of the invention  
XX  
SQ Sequence 557 AA;  
  
Query Match 0.9%; Score 9; DB 5; Length 557;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 944 SEKSSSES 952  
Db 18 SEKSSSES 26  
  
RESULT 20  
AAG49069  
ID AAG49069 standard; protein; 581 AA.  
XX  
AC AAG49069;  
XX  
XX 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 62038.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
PN  
XX  
XX PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
XX  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-013256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 18-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.

```
PR 09-JUL-1999; 99US-0142220P.
PR 12-JUL-1999; 99US-0142277P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149168P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.

PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161408P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      0.9%; Score 9; DB 3; Length 581;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      929 AAAAEAEQLL 937
Db      123 AAAAEAEQLL 131
      |||||
RESULT 21
ADJ50969
ID ADJ50969 standard; protein; 583 AA.
AC ADJ50969;
XX
XX 06-MAY-2004 (first entry)
DT Human novel protein NOV12a.
DE
XX
XX Human; NOVX; autoimmune disease; Alzheimer's disease; stroke; allergy;
KW Parkinson's disease; Huntington's disease; multiple sclerosis; anxiety;
KW pain; diabetes; graft versus host disease; pancreatitis; obesity; ulcer;
KW anaemia; cancer; viral infection; bacterial infection;
KW parasitic infection.
XX
XX Homo sapiens.
XX
XX US2004030096-A1.
XX
XX 12-FEB-2004.
XX
XX 01-AUG-2002; 2002US-00210281.
XX
XX 02-AUG-2001; 2001US-0309501P.
XX 03-AUG-2001; 2001US-0310291P.
XX 08-AUG-2001; 2001US-0310951P.
```



PR 09-AUG-2001; 2001US-0311292P.  
PR 13-AUG-2001; 2001US-0311979P.  
PR 14-AUG-2001; 2001US-03112203P.  
PR 17-AUG-2001; 2001US-03112031P.  
PR 20-AUG-2001; 2001US-03113643P.  
PR 20-AUG-2001; 2001US-03113702P.  
PR 21-AUG-2001; 2001US-03114031P.  
PR 23-AUG-2001; 2001US-03114466P.  
PR 28-AUG-2001; 2001US-03114039P.  
PR 29-AUG-2001; 2001US-03115853P.  
PR 05-MAR-2002; 2002US-0361775P.  
PR 05-MAR-2002; 2002US-0361832P.  
XX (GORM/) GORMAN L.  
PA (ZERH/) ZERHUSEN B D.  
PA (EDIN/) EDINGER S R.  
PA (PADI/) PADIGARU M.  
PA (GUOX/) GUO X.  
PA (KEKU/) KEKUDA R.  
PA (ZHON/) ZHONG M.  
PA (PATI/) PATTURAJAN M.  
PA (MILL/) MILLER C E.  
PA (JIWW/) JI W.  
PA (PENA/) PENNA C E A.  
PA (BURG/) BURGESS C E.  
PA (SCIO/) SCIORE P.  
PA (STON/) STONE D J.  
PA (TAUF/) TAUPIER R J.  
PA (CASM/) CASMAN S J.  
PA (ROTH/) ROTHENBERG M E.  
PA (MALY/) MALYANKAR U M.  
PA (BOLD/) BOLDOG F L.  
XX  
PI Gorman L, Zerkhusen BD, Edinger SR, Padigaru M, Guo X, Kekuda R;  
PI Zhong M, Patturajan M, Miller CE, Ji W, Pena CE, Burgess CE;  
PI Sciore P, Stone DJ, Taupier RJ, Casman SJ, Rothenberg ME;  
PI Malyankar UM, Boldog FL;  
XX  
DR WPI, 2004-168942/16.  
DR N-PSDB; ADJ50968.  
XX  
PT New NOVX polypeptides and polynucleotides, useful in diagnosing, treating  
PT or preventing diseases or conditions, e.g. autoimmune disease,  
PT Alzheimer's disease, diabetes, graft versus host disease, cancer or viral  
PT or bacterial infections.  
XX  
PS Claim 2; SEQ ID NO 34; 342pp; English.  
XX  
CC The invention relates to an isolated NOVX polypeptide (of 44 disclosed)  
CC comprising its mature form, a sequence having at least 95% sequence  
CC identity to NOVX or a sequence comprising one or more conservative  
CC substitutions in the amino acid sequence of NOVX. Also included are a  
CC composition comprising NOVX and a carrier, a kit comprising, in one or  
CC more containers, the composition, a method of identifying an agent that  
CC binds to NOVX, a method for identifying a potential therapeutic agent for  
CC use in treatment of a pathology related to aberrant expression or  
CC aberrant physiological interactions of NOVX, a method for screening for a  
CC modulator of activity of or of latency or predisposition to a pathology  
CC associated with NOVX, a method for modulating the activity of NOVX, a  
CC method of treating or preventing a pathology associated with NOVX or a  
CC pathological state in a mammal, an isolated nucleic acid molecule  
CC encoding a NOVX protein, a vector comprising the nucleic acid molecule,  
CC a cell comprising the vector, an antibody that immunospecifically binds  
CC to NOVX, a method for determining the presence or amount of NOVX or the  
CC nucleic acid molecule in a sample, a method for determining the presence  
CC of or predisposition to a disease associated with altered levels of  
CC expression of NOVX or the nucleic acid molecule in a first mammalian  
CC subject and a method of producing NOVX (comprising culturing the cell  
CC under conditions that lead to expression of the polypeptide). NOVX is  
CC useful in the manufacture of a medicament for treating a syndrome  
CC associated with a human disease associated with NOVX. The polypeptides  
CC and nucleic acid molecules are useful in diagnosing, treating or  
CC preventing diseases or conditions, e.g. autoimmune disease, Alzheimer's

CC disease, stroke, allergies, Parkinson's disease, Huntington's disease,  
CC multiple sclerosis, anxiety, pain, diabetes, graft versus host disease,  
CC pancreatitis, obesity, ulcers, anemia, cancer, viral or bacterial and  
CC parasitic infections (many more diseases and disorders are listed in the  
CC specification). The present sequence represents a NOVX protein.  
XX  
SQ Sequence 583 AA;

Query Match 0.9%; Score 9; DB 8; Length 583;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSSSES 952  
Db 18 SEKSSSSSES 26

## RESULT 22

AAG49068  
ID AAG49068 standard; protein; 590 AA.

XX  
AC AAG49068;

XX  
DT 18-OCT-2000 (first entry)

XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62037.

XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

XX  
PN EP1033405-A2.

XX  
PD 06-SEP-2000.

XX  
PF 25-FEB-2000; 2000EP-00301439.

XX  
PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126284P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 28-APR-1999; 99US-0130891P.

PR 30-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 04-MAY-1999; 99US-0132407P.

PR 06-MAY-1999; 99US-0132484P.

PR 06-MAY-1999; 99US-0132485P.

PR 07-MAY-1999; 99US-0132487P.

PR 11-MAY-1999; 99US-0132863P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 18-MAY-1999; 99US-0134370P.

PR 19-MAY-1999; 99US-0134768P.

PR 20-MAY-1999; 99US-0134941P.

PR 21-MAY-1999; 99US-0135124P.

PR 24-MAY-1999; 99US-0135353P.

PR 25-MAY-1999; 99US-0135629P.

PR 27-MAY-1999; 99US-0136021P.

PR 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155599P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157533P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0158293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

```
Query Match          0.9%; Score 9; DB 3; Length 590;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  929 AAAREQLL 937
DB  132 AAAREQLL 140
      |||||
      |||||

RESULT 23
ID  ABB93671 standard; protein; 590 AA.
XX  ABB93671;
AC  ABB93671;
XX  31-MAY-2002 (first entry)
XX  Herbicidally active polypeptide SEQ ID NO 2882.
XX  Herbicidal; plant; agriculture; herbicide.
XX  Arabidopsis thaliana.
XX  WO200210210-A2.
XX  07-FEB-2002.
XX  28-AUG-2001; 2001WO-EP009892.
XX  28-AUG-2001; 2001WO-EP009892.
XX  (FARB ) BAYER AG.
XX  Tietjen K, Weidner M;
XX  WPI; 2002-269010/31.
XX  Identifying plant target proteins for herbicidally active compounds,
XX  comprising aligning and comparing nucleic acid or amino acid sequences
XX  from plant with nucleic acid or amino acid sequences from non-plant
XX  organisms.
XX  Claim 5; SEQ ID NO 2882; 261pp + Sequence Listing; English.
XX  The invention relates to identifying target proteins (ABB90790-ABB94016)
XX  for herbicidally active compounds, comprising aligning and comparing
XX  nucleic acid or amino acid sequences from plant with nucleic acid or
XX  amino acid sequences from non-plant organisms using suitable search
XX  parameters, where plant sequences having an E-value greater by a factor
XX  of 3 than the E-value of most similar non-plant sequences are selected.
XX  The polypeptides or nucleic acids encoding them are useful for
XX  identifying modulators. The identified modulators are useful as
XX  herbicides
XX  Sequence 590 AA;

Query Match          0.9%; Score 9; DB 5; Length 590;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  929 AAAREQLL 937
DB  132 AAAREQLL 140
      |||||
      |||||

RESULT 24
ID  ABM85378 standard; protein; 608 AA.
XX  ABM85378;
AC  ABM85378;
XX  18-NOV-2004 (first entry)

Query Match          0.9%; Score 9; DB 7; Length 608;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  944 SEKSSSSSES 952
DB  36 SEKSSSSSES 44
      |||||
      |||||

RESULT 25
ID  AAG49067 standard; protein; 611 AA.
XX  AAG49067;
AC  AAG49067;
XX  18-OCT-2000 (first entry)
XX  Arabidopsis thaliana protein fragment SEQ ID NO: 62036.
XX  Protein identification; signal transduction pathway; metabolic pathway;
XX  hybridisation assay; genetic mapping; gene expression control; promoter;
XX  termination sequence.
XX  Arabidopsis thaliana.
XX  EP1033405-A2.
XX  06-SEP-2000.
XX  25-FEB-2000; 2000EP-00301439.

XX  Mouse protein sequence mCP21067.
XX  Cytostatic; carcinoma; lymphoma; cancer; murine.
XX  Mus musculus.
XX  WO20003073826-A2.
XX  12-SEP-2003.
XX  28-FEB-2003; 2003WO-US006235.
XX  01-MAR-2002; 2002US-00087192.
XX  (SAGR-) SAGRES DISCOVERY.
XX  Morris DW;
XX  WPI; 2003-328604/31.
XX  Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX  comprises a nucleotide sequence.
XX  Claim 5; SEQ ID NO 549; Opp; English.
XX  The present invention relates to novel DNA and protein sequences which
XX  are associated with carcinomas. The sequences are useful for: (i) for
XX  screening drug candidates; (ii) for screening of bioactive agent capable
XX  of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX  a bioactive agent capable of modulating the activity of CAP; (iv) for
XX  evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX  carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX  carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
XX  (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX  determining Carcinoma Associated (CA) gene copy number. In addition, the
XX  CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX  carcinoma including lymphoma. The present sequence is one such CAP. Note:
XX  This patent is an equivalent to basic patent US2002182586A1, for which no
XX  sequence data was published
XX  Sequence 608 AA;
```

XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 20-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.

PR 29-SEP-1999; 99US-0156596P.  
 PR 04-OCT-1999; 99US-0157117P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157865P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 22-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161992P.  
 PR 28-OCT-1999; 99US-0161993P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 0.9%; Score 9; DB 3; Length 611;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 929 AAAAEQQL 937  
 |||||  
 Db 153 AAAAEQQL 161

RESULT 26  
 AAW93890  
 ID AAW93890 standard; protein; 692 AA.

XX AC AAW93890;

XX DT 25-JUN-1999 (first entry)

XX DE Human HG38 protein.

XX KW HG38; human; G-protein coupled glycoprotein hormone receptor; brain;  
 endocrine system; skeletal muscle; placenta; development;  
 XX KW receptor activity modulator.

XX OS Homo sapiens.

XX PN WO9915660-A1.

XX PD 01-APR-1999.

XX PF 24-SEP-1998; 98WO-US019979.

XX PR 24-SEP-1997; 97US-0059863P.

XX PA (MERI) MERCK & CO INC.

XX PI Liu Q, Bailey WJ, McDonald TP;

XX

DR WPI; 1999-254711/21.  
 DR N-PSDB; AAX23981.  
 XX Human G-protein coupled glycoprotein hormone receptor HG38.  
 XX PS Disclosure; Fig 3A-E; 74pp; English.  
 XX CC This invention describes a novel human G-protein coupled glycoprotein  
 hormone receptor, HG38. Glycoprotein hormone receptors are important in  
 the endocrine system and HG38 may be involved in development and function  
 of the skeletal muscle, spinal cord, placenta and to a lesser extent, the  
 brain. The transgenic animal may be useful for studying tissue and  
 temporal specific expression or activity of the HG38 receptor, as well as  
 for studying the ability of a variety of compounds to act as modulators  
 of HG38 receptor activity  
 XX SQ Sequence 692 AA;

Query Match 0.9%; Score 9; DB 2; Length 692;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600  
 |||||  
 Db 174 LTEIPVQAF 182

RESULT 27

ADB80464

ID ADB80464 standard; protein; 883 AA.

XX AC ADB80464;

XX DT 04-DEC-2003 (first entry)

XX DE Ovarian cancer-associated protein #14.

XX KW cytostatic; gene therapy; vaccine; ovarian cancer; diagnosis;  
 post-operative chemotherapy; radiation therapy; tumour prognosis;  
 pre-cancerous lesion detection.

XX OS Homo sapiens.

XX PN WO2002102235-A2.

XX PD 27-DEC-2002.

XX PF 18-JUN-2002; 2002WO-US019297.

XX PR 18-JUN-2001; 2001US-0299234P.

XX PR 27-AUG-2001; 2001US-0315287P.

XX PR 05-SEP-2001; 2001US-0317544P.

XX PR 13-NOV-2001; 2001US-0350666P.

XX PR 12-APR-2002; 2002US-0372246P.

XX PA (EOSB-) EOS BIOTECHNOLOGY INC.

XX PI Mack DH, Gish KC;

XX DR WPI; 2003-167431/16.

XX DR N-PSDB; ADB80463.

XX PT Detecting an ovarian cancer-associated transcript in a cell from a  
 patient, comprises contacting a biological sample from the patient with a  
 polynucleotide that hybridizes to an ovarian cancer gene.

XX PS Claim 13; Page 284; 332pp; English.

XX CC The invention relates to a method of detecting an ovarian cancer-  
 associated transcript in a cell from a patient, by contacting a  
 biological sample from the patient with a polynucleotide that selectively  
 hybridizes to a sequence at least 80% identical to any of one of 80  
 nucleic acid sequences given in the specification. The method is useful

CC in diagnosing ovarian cancer and in identifying and using agents and/or  
 CC targets that inhibit ovarian cancer. The nucleic acid molecule,  
 CC polypeptide and the antibody may also be used in detecting ovarian  
 CC cancers, monitoring and early detection of relapse following treatment,  
 CC monitoring response to therapy, selecting patients for post-operative  
 CC chemotherapy or radiation therapy, in selecting mode of therapy,  
 CC determining tumour prognosis, early detection of pre-cancerous lesions,  
 CC and as vaccines. This sequence corresponds to one of the proteins used  
 CC for the detection method of the invention.

XX SQ Sequence 883 AA;

Query Match 0.9%; Score 9; DB 7; Length 883;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600  
 |||||  
 Db 174 LTEIPVQAF 182

RESULT 28

ADN40012  
 ID ADN40012 standard; protein; 883 AA.

XX AC ADN40012;

XX DT 17-JUN-2004 (first entry)

XX DE Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO:C382.

XX KW Human; differential expression; cancer; angiogenic disorder;  
 KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
 KW inflammatory disease; autoimmune disease;  
 KW retinal neovascularisation syndrome; scarring; uterine fibroid;  
 KW detection; diagnosis; prognosis; drug screening; drug targeting;  
 KW wound healing; contraception; cytostatic; cardiant; immunomodulatory;  
 KW vulnery; gene therapy; vaccine.

XX OS Homo sapiens.

XX PN WO2003042661-A2.

XX PD 22-MAY-2003.

XX PF 13-NOV-2002; 2002WO-US036810.

XX PR 13-NOV-2001; 2001US-0350666P.

XX PR 21-NOV-2001; 2001US-0332464P.

XX PR 29-NOV-2001; 2001US-0334393P.

XX PR 03-DEC-2001; 2001US-0335394P.

XX PR 14-DEC-2001; 2001US-0340376P.

XX PR 08-JAN-2002; 2002US-0347211P.

XX PR 10-JAN-2002; 2002US-0347349P.

XX PR 08-FEB-2002; 2002US-0355250P.

XX PR 13-FEB-2002; 2002US-0356714P.

XX PR 20-FEB-2002; 2002US-0359077P.

XX PR 29-MAR-2002; 2002US-0368809P.

XX PR 04-APR-2002; 2002US-0370110P.

XX PR 12-APR-2002; 2002US-0372246P.

XX PR 05-JUN-2002; 2002US-0386614P.

XX PR 16-JUL-2002; 2002US-0396839P.

XX PR 22-JUL-2002; 2002US-0397775P.

XX PR 22-JUL-2002; 2002US-0397845P.

XX PR 09-SEP-2002; 2002US-0409450P.

XX PA (EOSB-) EOS BIOTECHNOLOGY INC.

XX PI Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;  
 PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;

XX WPI; 2003-468649/44.

XX N-PSDB; ADN39795.

DR

XX Determining the presence or absence of a pathological cell in a patient,  
 PT useful for diagnosing, prognosing or treating cancer, comprises detecting  
 PT a nucleic acid in a biological sample.

XX PS Claim 12; SEQ ID NO C382; 1385pp; English.

XX The invention relates to nucleic acids and proteins (ADN3863-ADN40064)  
 CC whose expression is upregulated or downregulated in specific cancers or  
 CC other diseases such as angiogenic or fibrotic disorders, and to methods  
 CC of determining the presence or absence of a pathological cell in a  
 CC patient by detecting a nucleic acid at least 80% identical to those of  
 CC the invention or by detecting a polypeptide of the invention. The  
 CC invention also relates to expression vectors and host cells comprising a  
 CC nucleic acid of the invention; antibodies which specifically bind a  
 CC polypeptide of the invention; use of such antibodies for drug targeting;  
 CC and methods of screening for modulators of activity or expression of the  
 CC polypeptides and nucleic acids. The nucleic acids, polypeptides,  
 CC antibodies and methods are useful for diagnosing, prognosing and treating  
 CC cancer and other conditions such as psoriasis, ischaemia, heart disease,  
 CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
 CC neovascularisation syndromes, scarring and uterine fibroids. They may  
 CC also be useful in wound healing and in contraception. The present  
 CC sequence represents a polypeptide of the invention.

XX SQ Sequence 883 AA;

Query Match 0.9%; Score 9; DB 7; Length 883;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600  
 |||||  
 Db 174 LTEIPVQAF 182

RESULT 29

ADN39166

ID ADN39166 standard; protein; 883 AA.

XX AC ADN39166;

XX DT 17-JUN-2004 (first entry)

XX DE Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO:484.

XX KW Human; differential expression; cancer; angiogenic disorder;  
 KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
 KW inflammatory disease; autoimmune disease;  
 KW retinal neovascularisation syndrome; scarring; uterine fibroid;  
 KW detection; diagnosis; prognosis; drug screening; drug targeting;  
 KW wound healing; contraception; cytostatic; cardiant; immunomodulatory;  
 KW vulnery; gene therapy; vaccine.

XX OS Homo sapiens.

XX PN WO2003042661-A2.

XX PD 22-MAY-2003.

XX PF 13-NOV-2002; 2002WO-US036810.

XX PR 13-NOV-2001; 2001US-0350666P.

XX PR 21-NOV-2001; 2001US-0332464P.

XX PR 29-NOV-2001; 2001US-0334393P.

XX PR 03-DEC-2001; 2001US-0335394P.

XX PR 14-DEC-2001; 2001US-0340376P.

XX PR 08-JAN-2002; 2002US-0347211P.

XX PR 10-JAN-2002; 2002US-0347349P.

XX PR 08-FEB-2002; 2002US-0355250P.

XX PR 13-FEB-2002; 2002US-0356714P.

XX PR 20-FEB-2002; 2002US-0359077P.

XX PR 29-MAR-2002; 2002US-0368809P.





CC and AAA30775-A30779). The mutant proteins of the invention contain a  
 CC mutation in a portion of the protein comprising intracellular loop 3  
 CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,  
 CC is substituted for an endogenous residue in IC3 at a position 16 amino  
 CC acids N-terminal of an endogenous proline in TM6 to form a sequence X-  
 CC (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or  
 CC Ala, and is preferably Lys. When the endogenous residue at this position  
 CC is Lys, this residue is replaced by His, Arg or preferably Ala. The 15  
 CC amino acid stretch between the substituted amino acid and the Pro may be  
 CC endogenous, non-endogenous, or a mixture of endogenous and non-endogenous  
 CC residues. The constitutively active GPCRs are useful for identifying  
 CC antagonists, agonists and partial agonists for use as pharmaceutical  
 CC agents. The mutant proteins are also useful in research settings for  
 CC elucidating the roles of the receptors in normal and diseased conditions.  
 CC Antagonists for a particular GPCR are useful for treating diseases and  
 CC disorders associated with that receptor. Because the novel mutant GPCRs  
 CC are constitutively active, they can be used directly for screening of  
 CC compounds without the need for endogenous ligands. The present sequence  
 CC represents a human wild-type GPCR referred to in an exemplification of  
 CC the invention  
 CC  
 XX

XX Sequence 907 AA;

Query Match 0.9%; Score 9; DB 3; Length 907;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
 |||||  
 Db 174 LTEIPVQAF 182

RESULT 32

AAAY90687  
 ID AAAY90687 standard; protein; 907 AA.

XX AC AAAY90687;

XX DT 21-AUG-2000 (first entry)

XX DE Human mutant G protein-coupled receptor HG38 (V765K).

XX G protein-coupled receptor; GPCR; constitutively active;  
 KW intracellular loop 3; transmembrane domain 6; drug screening; agonist;  
 KW antagonist; mutant; mutein.

XX OS Homo sapiens.  
 OS Synthetic.

XX WO2000022129-A1.

XX PD 20-APR-2000.

XX PF 12-OCT-1999; 99WO-US023938.

XX PR 13-OCT-1998; 98US-00170496.

XX PA (AREN-) ARENA PHARM INC.

XX PI Behan DP, Chalmers DT, Liaw CW;

XX DR WPI; 2000-329165/28.

XX DR N-PSDB; AAA30779.

XX Non-endogenous constitutively activated human G protein-coupled  
 PT receptors, useful for identifying agonists for use as pharmaceutical  
 PT agents.

XX Example 2; Page 332-335; 341pp; English.

CC The invention relates to constitutively active, non-endogenous versions  
 CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAAY90643-  
 CC AAAY90677 and AAAY90683-Y90687), and to DNA encoding them (AAA30709-A30743

CC and AAA30775-A30779). The mutant proteins of the invention contain a  
 CC mutation in a portion of the protein comprising intracellular loop 3  
 CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,  
 CC is substituted for an endogenous residue in IC3 at a position 16 amino  
 CC acids N-terminal of an endogenous proline in TM6 to form a sequence X-  
 CC (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or  
 CC Ala, and is preferably Lys. When the endogenous residue at this position  
 CC is Lys, this residue is replaced by His, Arg or preferably Ala. The 15  
 CC amino acid stretch between the substituted amino acid and the Pro may be  
 CC endogenous, non-endogenous, or a mixture of endogenous and non-endogenous  
 CC residues. The constitutively active GPCRs are useful for identifying  
 CC antagonists, agonists and partial agonists for use as pharmaceutical  
 CC agents. The mutant proteins are also useful in research settings for  
 CC elucidating the roles of the receptors in normal and diseased conditions.  
 CC Antagonists for a particular GPCR are useful for treating diseases and  
 CC disorders associated with that receptor. Because the novel mutant GPCRs  
 CC are constitutively active, they can be used directly for screening of  
 CC compounds without the need for endogenous ligands. Sequences AAAY90643-  
 CC AAAY90677 and AAAY90683-Y90687 the mutant human GPCRs of the invention  
 XX

XX Sequence 907 AA;

Query Match 0.9%; Score 9; DB 3; Length 907;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
 |||||  
 Db 174 LTEIPVQAF 182

RESULT 33

ABP81968  
 ID ABP81968 standard; protein; 907 AA.

XX AC ABP81968;

XX DT 04-MAR-2003 (first entry)

XX DE Human G protein-coupled receptor GPR49 protein SEQ ID NO:422.

XX G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
 KW G protein-coupled receptor modulator; antibody; immune-related disease;  
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
 KW immunological-related cell proliferative disease; autoimmune disease;  
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
 KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;  
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;  
 KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  
 KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  
 KW ulcer.

XX OS Homo sapiens.

XX PN WO200261087-A2.

XX PD 08-AUG-2002.

XX PF 19-DEC-2001; 2001WO-US050107.

XX PR 19-DEC-2000; 2000US-0257144P.

XX PA (LIFE-) LIFESPAN BIOSCIENCES INC.

XX PI Burmer GC, Roush CL, Brown JP;

XX DR WPI; 2003-046718/04.

XX DR N-PSDB; ABZ42816.

XX New isolated antigenic peptides e.g., for G protein-coupled receptors  
 PT (GPCR), useful for diagnosing and designing drugs for treating conditions  
 PT in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or

PT autoimmune diseases.  
 XX Disclosure; Fig 1; 523pp; English.  
 PS  
 XX  
 CC The present invention describes antigenic peptides (I) comprising: (a)  
 CC any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
 CC acids. Also described: (1) an assay for the detection of a particular G  
 CC protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
 CC and (2) an isolated antibody having high specificity and high affinity or  
 CC avidity for a particular GPCR. (I) can be used as GPCR modulators and in  
 CC gene therapy. The antigenic peptides for GPCRs are useful in detecting an  
 CC antibody against a particular GPCR, and in the production of specific  
 CC antibodies. The peptides and antibodies are also useful for detecting the  
 CC presence or absence of corresponding GPCRs. The antigenic peptides for  
 CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
 CC treating immune-related diseases, growth-related diseases, cell  
 CC regeneration-related disease, immunological-related cell proliferative  
 CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
 CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
 CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
 CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
 CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
 CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
 CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
 CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  
 CC any other disorder in which GPCRs are involved. The antibodies may be  
 CC used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode  
 CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
 CC exemplification of the present invention  
 XX  
 XX Sequence 907 AA;  
 SQ

Query Match 0.9%; Score 9; DB 6; Length 907;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
 |||||  
 DB 174 LTEIPVQAF 182

RESULT 34  
 ABO06467  
 ID ABO06467 standard; protein; 907 AA.  
 XX  
 AC ABO06467;  
 XX  
 DT 13-AUG-2003 (first entry)  
 XX  
 DE Human G-protein coupled receptor HG38.  
 XX  
 KW G-protein coupled receptor; GPCR; caudate nucleus related disorder;  
 KW neoplastic disorder; leukaemia; breast cancer; immune disorder; AIDS;  
 KW rheumatoid arthritis; neuronal disorder; Alzheimer's disease; receptor;  
 KW Parkinson's disease; respiratory disorder; bronchopulmonary disease;  
 KW pancreatic's syndrome; ovarian disorder; ovarian carcinoma; colon disease;  
 KW Meig's syndrome; breast disease; thalamus related disorder; gene therapy;  
 KW amygdala related disorder; corpus callosum related disorder;  
 KW hippocampus related disorder.  
 OS Homo sapiens.  
 XX  
 XX US2003027323-A1.  
 PN 16-JAN-2001; 2001US-0261781P.  
 PR 19-JUL-2001; 2001US-030605P.  
 PR 03-AUG-2001; 2001US-0310436P.  
 XX

PA (FEDE/) FEDER J N.  
 PA (MINT/) MINTIER G.  
 PA (RAMA/) RAMANATHAN C S.  
 PA (HAWK/) HAWKEN D R.  
 XX  
 XX Feder JN, Mintier G, Ramanathan CS, Hawken DR;  
 PI WPI; 2003-466147/44.  
 XX  
 DR WPI; 2003-466147/44.  
 XX  
 PT New G-protein coupled receptor polypeptides, designated as HGPRBMY5,  
 PT useful for preventing, treating or ameliorating a medical condition  
 PT related to the colon, breast, ovaries or immune system.  
 XX  
 PS Disclosure; Fig 8; 93pp; English.  
 XX  
 CC The invention relates to an isolated HGPRBMY5 polypeptide. The  
 CC polypeptides, polynucleotides and methods are useful for preventing,  
 CC treating or ameliorating a medical condition such as a neoplastic  
 CC disorder e.g. leukaemia and breast cancer; immune disorder e.g. AIDS and  
 CC rheumatoid arthritis; neuronal disorder e.g. Alzheimer's disease and  
 CC Parkinson's disease; respiratory disorder e.g. bronchopulmonary disease  
 CC and pancreatic's syndrome; ovarian disorder e.g. ovarian carcinoma and  
 CC Meig's syndrome; colon disease; breast disease; thalamus related disorder  
 CC ; amygdala related disorder; corpus callosum related disorder; caudate  
 CC nucleus related disorder; hippocampus related disorder by administering  
 CC the GPCR polypeptide or its homologue. The present sequence represents  
 CC the amino acid sequence of a G-protein coupled receptor used to show  
 CC homology with the human G-protein coupled receptor, HGPRBMY  
 XX  
 XX Sequence 907 AA;  
 SQ

Query Match 0.9%; Score 9; DB 6; Length 907;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
 |||||  
 DB 174 LTEIPVQAF 182

RESULT 35  
 ADC22797  
 ID ADC22797 standard; protein; 907 AA.  
 XX  
 AC ADC22797;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE Human G protein-coupled receptor (GPCR) polypeptide #79.  
 XX  
 KW Human, G protein-coupled receptor; GPCR; transmembrane-6 region; TM6;  
 KW intracellular-3 region; IC3; receptor.  
 OS Homo sapiens.  
 XX  
 PN US5555339-B1.  
 XX  
 PD 29-APR-2003.  
 XX  
 XX 13-OCT-1998; 98US-00170496.  
 PF  
 XX  
 PR 14-APR-1997; 97US-00839449.  
 PR 14-APR-1998; 98US-00060188.  
 PR 26-JUN-1998; 98US-0090783P.  
 PR 07-AUG-1998; 98US-0095677P.  
 XX  
 XX (AREN-) ARENA PHARM INC.  
 PA  
 XX Liaw CW, Behan DP, Chalmers DT;  
 PI WPI; 2003-742861/70.  
 XX  
 DR N-PSDB; ADC22796.  
 XX

PT Creating a constitutively active version of an endogenous human G protein  
PT coupled receptor (GPCR) comprises substituting a specific amino acid in  
PT the transmembrane-6 region with a different amino acid, and testing for  
PT constitutive activity.

PS Example 2; SEQ ID NO 278; 221pp; English.

XX The invention relates to a method for treating a non-endogenous,  
CC constitutively active version of an endogenous human G protein-coupled  
CC receptor (GPCR) that has a transmembrane-6 (TM6) region and an  
CC intracellular-3 (IC3) region, by substituting a specific amino acid in  
CC the TM6 region with a different amino acid, and testing for constitutive  
CC activity. The method is useful for creating a constitutively active  
CC version of an endogenous human GPCR that comprises a transmembrane 6  
CC region and an intracellular loop 3 region. The altered human GPCR  
CC polypeptides are useful for screening test compounds for identification  
CC of inverse agonists or partial agonists of GPCR polypeptides, which may  
CC have therapeutic uses. The altered GPCRs may also be used in vivo or in  
CC vitro in biological research. A nucleic acid encoding the altered GPCR  
CC may be used to create a transgenic animal expressing the altered GPCR.  
CC The method allows screening for compounds that modulate the activity of a  
CC human G protein-coupled receptor without the need for provision of a  
CC ligand for the receptor. This is particularly useful in allowing  
CC screening of compounds against orphan receptors for which no ligand is  
CC currently known. This sequence represents a human GPCR polypeptide of the  
CC invention.

SQ Sequence 907 AA;

Query Match 0.9%; Score 9; DB 7; Length 907;  
Best Local Similarity 100.0%; Pred. No. 45;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
Db 174 LTEIPVQAF 182  
|||||||

RESULT 36

ADC22783

ID ADC22783 standard; protein; 907 AA.

XX AC ADC22783;

XX DT 18-DEC-2003 (first entry)

XX DE Human G protein-coupled receptor (GPCR) polypeptide #39.

XX KW Human; G protein-coupled receptor; GPCR; transmembrane-6 region; TM6;  
KW intracellular-3 region; IC3; receptor.

XX OS Homo sapiens.

XX PN US6555339-B1.

XX PD 29-APR-2003.

XX PF 13-OCT-1998; 98US-00170496.

XX PR 14-APR-1997; 97US-00839449.

XX PR 14-APR-1998; 98US-00060188.

XX PR 26-JUN-1998; 98US-0090783P.

XX PR 07-AUG-1998; 98US-0095677P.

XX PA (AREN-) ARENA PHARM INC.

XX PI Liaw CW, Behan DE, Chalmers DT;

XX DR WPI; 2003-742861/70.

XX DR N-PSDB; ADC22782.

XX PT Creating a constitutively active version of an endogenous human G protein  
PT coupled receptor (GPCR) comprises substituting a specific amino acid in

PT the transmembrane-6 region with a different amino acid, and testing for  
PT constitutive activity.

PS Example 1; SEQ ID NO 264; 221pp; English.

XX The invention relates to a method for treating a non-endogenous,  
CC constitutively active version of an endogenous human G protein-coupled  
CC receptor (GPCR) that has a transmembrane-6 (TM6) region and an  
CC intracellular-3 (IC3) region, by substituting a specific amino acid in  
CC the TM6 region with a different amino acid, and testing for constitutive  
CC activity. The method is useful for creating a constitutively active  
CC version of an endogenous human GPCR that comprises a transmembrane 6  
CC region and an intracellular loop 3 region. The altered human GPCR  
CC polypeptides are useful for screening test compounds for identification  
CC of inverse agonists or partial agonists of GPCR polypeptides, which may  
CC have therapeutic uses. The altered GPCRs may also be used in vivo or in  
CC vitro in biological research. A nucleic acid encoding the altered GPCR  
CC may be used to create a transgenic animal expressing the altered GPCR.  
CC The method allows screening for compounds that modulate the activity of a  
CC human G protein-coupled receptor without the need for provision of a  
CC ligand for the receptor. This is particularly useful in allowing  
CC screening of compounds against orphan receptors for which no ligand is  
CC currently known. This sequence represents a human GPCR polypeptide of the  
CC invention.

SQ Sequence 907 AA;

Query Match 0.9%; Score 9; DB 7; Length 907;  
Best Local Similarity 100.0%; Pred. No. 45;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
Db 174 LTEIPVQAF 182  
|||||||

RESULT 37

ADE59150

ID ADE59150 standard; protein; 907 AA.

XX AC ADE59150;

XX DT 29-JAN-2004 (first entry)

XX DE Human Protein O75473, SEQ ID NO 5041.

XX KW Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.

XX OS Homo sapiens.

XX PN WO2003016475-A2.

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002WO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346382P.

XX PR 26-NOV-2001; 2001US-0333347P.

XX PA (GHEO) GEN HOSPITAL CORP.

XX PA (FARB) BAYER AG.

XX PI Woolf C, D'urso D, Befort K, Costigan M;

XX DR WPI; 2003-268312/26.

XX DR GENBANK; 075473.

XX PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.

PS Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat

CC or human polynucleotides or a polynucleotide which represents a fragment,

CC derivative or allelic variation of the nucleic acid sequence. Also

CC claimed are a vector comprising the novel polynucleotide, a host cell

CC comprising the vector, a method for identifying a nucleotide sequence

CC which is differentially regulated in an animal subjected to pain and a

CC kit to perform the method, an array, a method for identifying an agent

CC that increases or decreases the expression of the polynucleotide sequence

CC that is differentially expressed in neuronal tissue of a first animal

CC subjected to pain, a method for identifying a compound which regulates

CC the expression of a polynucleotide sequence which is differentially

CC expressed in an animal subjected to pain, a method for identifying a

CC compound that regulates the activity of one or more of the

CC polynucleotides, a method for producing a pharmaceutical composition, a

CC method for identifying a compound or small molecule that regulates the

CC activity in an animal of one or more of the polypeptides given in the

CC specification, a method for identifying a compound useful in treating

CC pain and a pharmaceutical composition comprising the one or more

CC polypeptides or their antibodies. The polynucleotide or the compound that

CC modulates its activity is useful for preparing a medicament for treating

CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene

CC therapy). The sequence presented is a human protein (shown in Table 2 of

CC the specification) which is differentially expressed during pain. Note:

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic form directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 907 AA;

QY 592 LTEIPVQAF 600

DB 174 LTEIPVQAF 182

RESULT 38

AD59153

ID ADE59153 standard; protein; 907 AA.

AC ADE59153;

XX 29-JAN-2004 (first entry)

DE Human Protein O75473, SEQ ID NO 5044.

XX Human; pain; neuronal tissue; gene therapy;

KW spinal segmental nerve injury; chronic constriction injury; CCI;

KW spared nerve injury; SNI; Chung.

XX Homo sapiens.

OS WO2003016475-A2.

PN 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

PF 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GEO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

PI WPI; 2003-268312/26.

DR

GENBANK; O75473.

DR New composition comprising two or more isolated polypeptides, useful for

XX preparing a medicament for treating pain in an animal.

PT Claim 1; Page; 1017pp; English.

PS The invention discloses a composition comprising two or more isolated rat

CC or human polynucleotides or a polynucleotide which represents a fragment,

CC derivative or allelic variation of the nucleic acid sequence. Also

CC claimed are a vector comprising the novel polynucleotide, a host cell

CC comprising the vector, a method for identifying a nucleotide sequence

CC which is differentially regulated in an animal subjected to pain and a

CC kit to perform the method, an array, a method for identifying an agent

CC that increases or decreases the expression of the polynucleotide sequence

CC that is differentially expressed in neuronal tissue of a first animal

CC subjected to pain, a method for identifying a compound which regulates

CC the expression of a polynucleotide sequence which is differentially

CC expressed in an animal subjected to pain, a method for identifying a

CC compound that regulates the activity of one or more of the

CC polynucleotides, a method for producing a pharmaceutical composition, a

CC method for identifying a compound or small molecule that regulates the

CC activity in an animal of one or more of the polypeptides given in the

CC specification, a method for identifying a compound useful in treating

CC pain and a pharmaceutical composition comprising the one or more

CC polypeptides or their antibodies. The polynucleotide or the compound that

CC modulates its activity is useful for preparing a medicament for treating

CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene

CC therapy). The sequence presented is a human protein (shown in Table 2 of

CC the specification) which is differentially expressed during pain. Note:

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic form directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 907 AA;

Query Match 0.9%; Score 9; DB 7; Length 907;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600

DB 174 LTEIPVQAF 182

RESULT 39

ADG42629

ID ADG42629 standard; protein; 907 AA.

AC ADG42629;

XX 26-FEB-2004 (first entry)

DE Human G protein-coupled receptor LGR5.

XX cytostatic; gene therapy; NOX-agonist; NOX-antagonist; pharmaceutical;

KW NOX-associated disorder; cancer; human; G protein coupled receptor LGR5.

OS Homo sapiens.

OS US2003204052-A1.

PN 30-OCT-2003.

PF 04-OCT-2001; 2001US-00970944.

PR 04-OCT-2000; 2000US-0237862P.

XX (HERR) HERRMANN J L.

PA (RAST) RASTELLI L.

PA (SHIM) SHIMKETS R A.

XX



Best Local Similarity 100.0%; Pred. No. 45;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600  
Db 174 LTEIPVQAF 182

RESULT 42  
ADH14256  
ID ADH14256 standard; protein; 907 AA.  
AC ADH14256;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Human HG38.  
XX  
KW human; non-endogenous; G protein-coupled receptor; GPCR; receptor.  
XX  
OS Homo sapiens.  
XX  
FN US2003105292-A1.  
XX  
PD 05-JUN-2003.  
XX  
PF 20-SEP-2002; 2002US-00251385.  
XX  
PR 26-JUN-1998; 98US-0090783P.  
PR 07-AUG-1998; 98US-0095677P.  
PR 13-OCT-1998; 98US-00170496.  
XX  
PA (LIAM/) LIAM C W.  
PA (BEHA/) BEHAN D P.  
PA (CHAL/) CHALMERS D T.  
XX  
PI Liaw CW, Behan DP, Chalmers DT;  
XX  
DR WPI; 2003-801247/75.  
DR N-PSDB; ADH14255.  
XX  
PT New constitutively active, non-endogenous version of an endogenous human  
PT G protein-coupled receptor for the identification of therapeutic  
PT compounds, such as agonists.  
XX  
PS Example 1; SEQ ID NO 264; 227pp; English.  
XX  
CC The invention relates to a constitutively active, non-endogenous version  
CC of an endogenous human G protein-coupled receptor (GPCR). The GPCR is  
CC used for screening therapeutic compounds as inverse agonists, agonists or  
CC partial agonists. The GPCR can be also be used to elucidate and  
CC understand the roles of GPCRs in normal and diseased humans. The GPCR  
CC need not be purified and isolated to be used to screen for therapeutic  
CC compounds. The utility of the GPCR as a research tool is enhanced because  
CC the role of a particular receptor can be understood before the endogenous  
CC ligand is identified. The present sequence is used in the exemplification  
XX of the present invention.

Query Match 0.9%; Score 9; DB 7; Length 907;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600  
Db 174 LTEIPVQAF 182

RESULT 43  
ADN40013  
ID ADN40013 standard; protein; 907 AA.  
XX

AC ADN40013;  
XX  
DT 17-JUN-2004 (first entry)  
XX  
DE Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO:C383.  
XX  
KW Human; differential expression; cancer; angiogenic disorder;  
KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
KW inflammatory disease; autoimmune disease;  
KW retinal neovascularisation syndrome; scarring; uterine fibroid;  
KW detection; diagnosis; prognosis; drug screening; drug targeting;  
KW wound healing; contraception; cytostatic; cardiac; immunomodulatory;  
KW vulnerary; gene therapy; vaccine.  
XX  
OS Homo sapiens.  
XX  
FN WO2003042661-A2.  
XX  
PD 22-MAY-2003.  
XX  
PF 13-NOV-2002; 2002WO-US036810.  
XX  
PR 13-NOV-2001; 2001US-0350666P.  
PR 21-NOV-2001; 2001US-0332464P.  
PR 29-NOV-2001; 2001US-0334393P.  
PR 03-DEC-2001; 2001US-0335394P.  
PR 14-DEC-2001; 2001US-0340376P.  
PR 08-JAN-2002; 2002US-0347211P.  
PR 10-JAN-2002; 2002US-0347349P.  
PR 08-FEB-2002; 2002US-0355250P.  
PR 13-FEB-2002; 2002US-0356714P.  
PR 20-FEB-2002; 2002US-0359077P.  
PR 09-MAR-2002; 2002US-0368809P.  
PR 04-APR-2002; 2002US-0370110P.  
PR 12-APR-2002; 2002US-0372246P.  
PR 05-JUN-2002; 2002US-0386614P.  
PR 16-JUL-2002; 2002US-0396839P.  
PR 22-JUL-2002; 2002US-0397775P.  
PR 22-JUL-2002; 2002US-0397845P.  
PR 09-SEP-2002; 2002US-0409450P.  
XX  
PA (EOSB-) EOS BIOTECHNOLOGY INC.  
XX  
PI Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;  
PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;  
XX  
DR WPI; 2003-468649/44.  
DR N-PSDB; ADN39796.  
XX  
PT Determining the presence or absence of a pathological cell in a patient,  
PT useful for diagnosing, prognosing or treating cancer, comprises detecting  
PT a nucleic acid in a biological sample.  
XX  
PS Claim 12; SEQ ID NO C383; 1385pp; English.  
XX  
CC The invention relates to nucleic acids and proteins (ADN38683-ADN40064)  
CC whose expression is upregulated or downregulated in specific cancers or  
CC other diseases such as angiogenic or fibrotic disorders, and to methods  
CC of determining the presence or absence of a pathological cell in a  
CC patient by detecting a nucleic acid at least 80% identical to those of  
CC the invention or by detecting a polypeptide of the invention. The  
CC invention also relates to expression vectors and host cells comprising a  
CC nucleic acid of the invention; antibodies which specifically bind a  
CC polypeptide of the invention; use of such antibodies for drug targeting;  
CC and methods of screening for modulators of activity or expression of the  
CC polypeptides and nucleic acids. The nucleic acids, polypeptides,  
CC antibodies and methods are useful for diagnosing, prognosing and treating  
CC cancer and other conditions such as psoriasis, ischaemia, heart disease,  
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
CC neovascularisation syndromes, scarring and uterine fibroids. They may  
CC also be useful in wound healing and in contraception. The present  
CC sequence represents a polypeptide of the invention.











CC encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2  
CC domain and selecting plants the with increased size or mass. The  
CC invention is useful to generate transgenic plants. The method is useful  
CC for increasing cell proliferation in a plant for altering organ mass,  
CC controlling fertility or enhancing asexual reproduction. The present  
CC sequence is Arabidopsis thaliana ANT mutant protein. This sequence is  
CC used in the exemplification of the invention.  
XX  
SQ Sequence 196 AA;

Query Match 0.8%; Score 8; DB 7; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTL 12  
|||||||  
DB 177 LKTVALTL 184

RESULT 52  
ADM26567  
ID ADM26567 standard; protein; 240 AA.

AC ADM26567;

XX 20-MAY-2004 (first entry)

XX Hyperthermophile Methanopyrus kandleri protein #1173.

XX hyperthermophile; protein stability enhancement;  
KW protein activity enhancement.

XX Methanopyrus kandleri.

XX WO2003076575-A2.

XX 18-SEP-2003.

XX 04-MAR-2003; 2003WO-US006664.

XX 04-MAR-2002; 2002US-0361742P.

XX 14-MAY-2002; 2002US-0380423P.

XX 16-SEP-2002; 2002US-0410974P.

XX (FIDE-) FIDELITY SYSTEMS INC.

XX (MALY/) MALYKH A.

XX Slesarev AI, Pavlov A, Pavlova N, Kozyavkin S;

XX WPI; 2003-748383/70.

XX N-PSDB; ADM27081.

XX New isolated nucleic acids encoding any of about 1700 Methanopyrus

XX kandleri proteins, and the encoded proteins, useful as a medicaments or  
XX as diagnostic agents.  
XX Claim 31; SEQ ID NO 1173; 1023pp; English.  
XX  
XX The invention comprises the amino acid sequence of proteins from the  
XX hyperthermophile Methanopyrus kandleri, the invention also comprises the  
XX complete genome from Methanopyrus kandleri. The Methanopyrus kandleri  
XX proteins of the invention are useful for enhancing the stability and/or  
XX activity of other proteins. The Methanopyrus kandleri genome is useful in  
XX a variety of diagnostic and analytical methods. The present amino acid  
XX sequence represents a Methanopyrus kandleri protein of the invention.

XX Sequence 240 AA;

Query Match 0.8%; Score 8; DB 7; Length 240;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 794 RLEKELDL 801

DB 44 RLEKELDL 51  
|||||||  
RESULT 53  
ADG88453  
ID ADG88453 standard; protein; 264 AA.

XX ADG88453;

XX 11-MAR-2004 (first entry)

XX Arabidopsis thaliana aintegumenta (ANT) mutant protein, ANT-AP2sw2.

XX Plant; Genetic engineering; cell proliferation; aintegumenta; ANT;

XX transgenic; transgenic plant; organ mass alteration; fertility;

XX asexual reproduction; mutant; mutain.

XX Synthetic.

XX Arabidopsis thaliana.

XX US2003159180-A1.

XX 21-AUG-2003.

XX 28-JAN-2002; 2002US-00059911.

XX 28-JAN-2002; 2002US-00059911.

XX (REGC ) UNIV CALIFORNIA OFFICE TECHNOLOGY.

XX Fischer RL, Mizukami Y;

XX WPI; 2003-787370/74.

XX N-PSDB; ADG8470.

XX Increasing cell proliferation in a plant by introducing into the plant an

XX expression cassette containing a plant promoter operably linked to a

XX nucleic acid encoding a modified ANT polypeptide and selecting plants

XX with increased mass.

XX Disclosure; SEQ ID NO 24; 51pp; English.

XX The present invention relates to plant genetic engineering. The invention  
XX particularly relates to a method of increasing cell proliferation in a  
XX plant. The method involves introducing into the plant an expression  
XX cassette containing a plant promoter operably linked to a nucleic acid  
XX encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2  
XX domain and selecting plants the with increased size or mass. The  
XX invention is useful to generate transgenic plants. The method is useful  
XX for increasing cell proliferation in a plant for altering organ mass,  
XX controlling fertility or enhancing asexual reproduction. The present  
XX sequence is Arabidopsis thaliana ANT mutant protein. This sequence is  
XX used in the exemplification of the invention.

XX Sequence 264 AA;

Query Match 0.8%; Score 8; DB 7; Length 264;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTL 12

DB 245 LKTVALTL 252

RESULT 54

ABB05713

ID ABB05713 standard; protein; 268 AA.

XX ABB05713;

XX 30-APR-2002 (first entry)

XX DE Human cell structure and motility protein clone tes3\_16b5.  
 XX KW Human; foetal brain; foetal kidney; melanoma; testis; amygdala;  
 XX KW gene therapy.  
 XX OS Homo sapiens.  
 XX PN WO200198454-A2.  
 XX PD 27-DEC-2001.  
 XX PF 25-APR-2001; 2001WO-IB002050.  
 XX PR 25-APR-2000; 2000US-0199380P.  
 XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.  
 XX PI Wiemann S;  
 XX XN WPI; 2002-055860/07.  
 XX DR N-PSDB; ABA93750.  
 XX PT Human cDNA sequences and clones derived from human fetal brain, fetal  
 PT kidney, melanoma, testis and amygdala cDNA libraries, useful in genetic  
 PT screening and therapy.  
 XX PS Claim 1; Page 315; 61lpp; English.  
 XX CC The present invention describes assemblages and computer readable media  
 CC comprising novel human cDNA sequences and clones derived from human  
 CC foetal brain, foetal kidney, melanoma, testis and amygdala cDNA  
 CC libraries. ABA93702 to ABA93766 represent human cDNA sequences from the  
 CC present invention which encode the proteins given in ABB05662 to  
 CC ABB05729. The human cDNA sequences and clones can be used in gene  
 CC therapy. The clones may be used in a variety of applications, for example  
 CC they may be used in profiling assays, for providing large arrays of human  
 CC genetic material for implementing large-scale screening strategies and  
 CC for treating diseases via gene therapy procedures  
 XX SQ Sequence 268 AA;  
 Query Match 0.8%; Score 8; DB 5; Length 268;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 793 KELEKELD 800  
 DB 83 KELEKELD 90  
 RESULT 55  
 ADG88454  
 ID ADG88454 standard; protein; 275 AA.  
 XX AC ADG88454;  
 XX DT 11-MAR-2004 (first entry)  
 XX DE Arabidopsis thaliana aintegumenta (ANT) mutant protein, ANTdf1.  
 XX KW Plant; genetic engineering; cell proliferation; aintegumenta; ANT;  
 KW transgenic; transgenic plant; organ mass alteration; fertility;  
 KW asexual reproduction; mutant; mutein.  
 XX OS Synthetic.  
 OS Arabidopsis thaliana.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 1..2 /note= "Encoded by ATGACTTCT"  
 FT Misc-difference 134 /note= "Encoded by GGA"

XX PN US2003159180-A1.  
 XX PD 21-AUG-2003.  
 XX PF 28-JAN-2002; 2002US-00059911.  
 XX PR 28-JAN-2002; 2002US-00059911.  
 XX PA (REGC ) UNIV CALIFORNIA OFFICE TECHNOLOGY.  
 XX PI Fischer RL, Mizukami Y;  
 XX XN WPI; 2003-787370/74.  
 XX DR N-PSDB; ADG88471.  
 XX PT Increasing cell proliferation in a plant by introducing into the plant an  
 PT expression cassette containing a plant promoter operably linked to a  
 PT nucleic acid encoding a modified ANT polypeptide and selecting plants  
 PT with increased mass.  
 XX PS Claim 4; SEQ ID NO 25; 5lpp; English.  
 XX CC The present invention relates to plant genetic engineering. The invention  
 CC particularly relates to a method of increasing cell proliferation in a  
 CC plant. The method involves introducing into the plant an expression  
 CC cassette containing a plant promoter operably linked to a nucleic acid  
 CC encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2  
 CC domain and selecting plants the with increased size or mass. The  
 CC invention is useful to generate transgenic plants. The method is useful  
 CC for increasing cell proliferation in a plant for altering organ mass,  
 CC controlling fertility or enhancing asexual reproduction. The present  
 CC sequence is Arabidopsis thaliana ANT mutant protein. This sequence is  
 CC used in the exemplification of the invention.  
 XX SQ Sequence 275 AA;  
 Query Match 0.8%; Score 8; DB 7; Length 275;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 LKTVALTL 12  
 DB 256 LKTVALTL 263  
 RESULT 56  
 ADG88440  
 ID ADG88440 standard; protein; 277 AA.  
 XX AC ADG88440;  
 XX DT 11-MAR-2004 (first entry)  
 XX DE Arabidopsis thaliana aintegumenta (ANT) mutant protein, ANTfDN3.  
 XX KW Plant; genetic engineering; cell proliferation; aintegumenta; ANT;  
 KW transgenic; transgenic plant; organ mass alteration; fertility;  
 KW asexual reproduction; mutant; mutein.  
 XX OS Synthetic.  
 OS Arabidopsis thaliana.  
 XX PN US2003159180-A1.  
 XX PD 21-AUG-2003.  
 XX PF 28-JAN-2002; 2002US-00059911.  
 XX PR 28-JAN-2002; 2002US-00059911.  
 XX PA (REGC ) UNIV CALIFORNIA OFFICE TECHNOLOGY.  
 XX



CC particularly relates to a method of increasing cell proliferation in a  
 CC plant. The method involves introducing into the plant an expression  
 CC cassette containing a plant promoter operably linked to a nucleic acid  
 CC encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2  
 CC domain and selecting plants the with increased size or mass. The  
 CC invention is useful to generate transgenic plants. The method is useful  
 CC for increasing cell proliferation in a plant for altering organ mass,  
 CC controlling fertility or enhancing asexual reproduction. The present  
 CC sequence is Arabidopsis thaliana ANT mutant protein. This sequence is  
 CC used in the exemplification of the invention.

XX Sequence 308 AA;

Query Match 0.8%; Score 8; DB 7; Length 308;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LKTVALTL 12

Db 289 LKTVALTL 296

RESULT 59

AAV03187  
 ID AAY03187 standard; protein; 319 AA.

XX  
 AC AAY03187;

XX  
 DT 16-JUN-1999 (first entry)

XX Rat Acid sensitive ion channel protein sequence.

XX Acid sensitive ion channel; rat; ASIC; pH mediated pain disorder;  
 KW ischaemia; Gene therapy; proton-gated ion channel.

XX Rattus sp.

XX WO9911784-A1.

XX  
 PD 11-MAR-1999.

XX  
 PF 28-AUG-1998; 98WO-GB002609.

XX  
 PR 29-AUG-1997; 97GB-00018365.

XX (UNLO ) UNIV COLLEGE LONDON.

XX Wood JN, England S, Akopian AN, Chen C;

XX  
 DR WPI; 1999-205188/17.

XX  
 DR N-PSDB; AAX28162.

XX Acid sensitive ion channel (ASIC) proteins - useful in gene therapy for  
 PT treatment of pH mediated pain disorders.

XX Claim 5; Page 51-53; 62pp; English.

XX This sequence represents an acid sensitive ion channel (ASIC) of the  
 CC invention, isolated from rat dorsal root ganglion. Antisense ASIC DNA or  
 CC RNA is useful in gene therapy for downgrading expression of ASIC protein,  
 CC for pH mediated pain disorders e.g. in ischaemia. The vectors are useful  
 CC for incorporating ASIC DNA or RNA for use in gene therapy. Proton-gated  
 CC ion channel agonists and antagonists are identified using cells  
 CC transfected with ASIC DNA by allowing interaction between the candidate  
 CC substance and ASIC protein in the membrane, and measuring interaction  
 CC and/or cell response. Partial agonists and antagonists can be identified  
 CC by their ability to block the response of the cell to present in a  
 CC solution of a given acid pH or any agonist. The hybridisation probes are  
 CC useful for screening libraries for ASIC DNA or RNA

XX Sequence 319 AA;

Query Match 0.8%; Score 8; DB 2; Length 319;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 QASSSIG 44

Db 306 QASSSIG 313

RESULT 60

AAV22200  
 ID AAY22200 standard; protein; 350 AA.

XX  
 AC AAY22200;

XX  
 DT 10-SEP-1999 (first entry)

XX S. frugiperda retinol dehydratase protein sequence.

XX Retinol dehydratase; enzyme; cell growth; cell proliferation; retinol;  
 KW 14-hydroxy-4,14-retro-retinol; 14-HRR; immune system; allergic reaction;  
 KW humoral immune response; graft versus host disease; psoriasis; therapy;  
 KW autoimmune disease.

XX Spodoptera frugiperda.

XX US5928931-A.

XX  
 PD 27-JUL-1999.

XX  
 PF 07-MAY-1997; 97US-00852481.

XX  
 PR 09-MAY-1996; 96US-0017178P.

XX (CORR ) CORNELL RES FOUND INC.

PA (SLOK ) SLOAN KETTERING INST CANCER RES.

XX Hammerling U, Gruen F, Buck J;

XX  
 DR WPI; 1999-429507/36.

XX  
 DR N-PSDB; AAX84491.

XX Nucleic acids encoding retinol dehydratase useful for antagonizing cell  
 PT growth and proliferation mediated by 14-hydroxy-4,14-retro-retinol or  
 PT retinol.

XX Claim 2; Fig 4; 23pp; English.

XX This sequence is the Spodoptera frugiperda retinol dehydratase of the  
 CC invention. The isolated nucleic acids may be used to transform host cells  
 CC for the recombinant production of retinol dehydratase enzymes. These  
 CC enzymes may be used to antagonise cell growth and proliferation mediated  
 CC by retinol and 14-hydroxy-4,14-retro-retinol (14-HRR) and prevent  
 CC uncontrolled cellular replication. Processes as diverse as reproduction,  
 CC growth, vision and pattern formation and differentiation during  
 CC embryogenesis are influenced by retinol and 14-HRR, and so may be  
 CC antagonised in this way. In particular, the immune system is influenced  
 CC by 14-HRR and so it may be modulated using retinol dehydratase enzymes.  
 CC Diseases characterised by humoral immune responses which may be treated  
 CC using the enzyme include graft versus host disease, psoriasis, allergic  
 CC reactions and autoimmune diseases. The enzyme may be recombinantly  
 CC produced both in fermentation culture and in vivo by transformation of  
 CC affected cells with the nucleic acids. The nucleic acid molecules may  
 CC also be used as probes in assays to detect and quantify expression of  
 CC retinol dehydratase

XX Sequence 350 AA;

Query Match 0.8%; Score 8; DB 2; Length 350;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 668 LADGEHFI 675

|||||||



```
Db      303 LADGEHFI 310
RESULT 61
ABU17197
ID      ABU17197 standard; protein; 356 AA.
XX
AC      ABU17197;
XX
DT      19-JUN-2003 (first entry)
XX
DE      Protein encoded by Prokaryotic essential gene #2724.
XX
KW      Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS      Acinetobacter baumannii.
XX
PN      WO200277183-A2.
XX
PD      03-OCT-2002.
XX
PF      21-MAR-2002; 2002WO-US009107.
XX
PR      21-MAR-2001; 2001US-00815242.
PR      06-SEP-2001; 2001US-00948993.
PR      25-OCT-2001; 2001US-0342923P.
PR      08-FEB-2002; 2002US-00072851.
PR      06-MAR-2002; 2002US-0362699P.
XX
PA      (ELIT-) ELITRA PHARM INC.
XX
PI      Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI      Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR      WPI; 2003-029926/02.
DR      N-PSDB; ACA21067.
XX
PT      New antisense nucleic acids, useful for identifying proteins or screening
PT      for homologous nucleic acids required for cellular proliferation to
PT      isolate candidate molecules for rational drug discovery programs.
XX
PS      Claim 25; SEQ ID NO 45121; 1766pp; English.
XX
CC      The invention relates to an isolated nucleic acid comprising any one of
CC      the 6213 antisense sequences given in the specification where expression
CC      of the nucleic acid inhibits proliferation of a cell. Also included are:
CC      (1) a vector comprising a promoter operably linked to the nucleic acid
CC      encoding a polypeptide whose expression is inhibited by the antisense
CC      nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC      polypeptide or its fragment whose expression is inhibited by the
CC      antisense nucleic acid; (4) an antibody capable of specifically binding
CC      the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC      proliferation or the activity of a gene in an operon required for
CC      proliferation; (7) identifying a compound that influences the activity of
CC      the gene product or that has an activity against a biological pathway
CC      required for proliferation, or that inhibits cellular proliferation; (8)
CC      identifying a gene required for cellular proliferation or the biological
CC      pathway in which a proliferation-required gene or its gene product lies
CC      or a gene on which the test compound that inhibits proliferation of an
CC      organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC      compound's activity; (11) a culture comprising strains in which the gene
CC      product is overexpressed or underexpressed; (12) determining the extent
CC      to which each of the strains is present in a culture or collection of
CC      strains; or (13) identifying the target of a compound that inhibits the
CC      proliferation of an organism. The antisense nucleic acids are useful for
CC      identifying proteins or screening for homologous nucleic acids required
CC      for cellular proliferation to isolate candidate molecules for rational
CC      drug discovery programs, or for screening homologous nucleic acids
CC      required for proliferation in cells other than S. aureus, S. typhimurium,
CC      K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC      the target prokaryotic essential genes. Note: The sequence data for this
CC      patent did not form part of the printed specification, but was obtained
CC      in electronic format directly from WIPO at
```

```
CC      ftp.wipo.int/pub/published_pct_sequences
XX
SQ      Sequence 356 AA;
Query Match      0.8%; Score 8; DB 6; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
QY      997 LSLITAIK 1004
      |||||
DB      156 LSLITAIK 163
      |||||
RESULT 62
ADA33080
ID      ADA33080 standard; protein; 364 AA.
XX
AC      ADA33080;
XX
DT      20-NOV-2003 (first entry)
XX
DE      Acinetobacter baumannii protein #241.
XX
KW      Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW      plant biocontrol agent.
XX
OS      Acinetobacter baumannii.
XX
PN      US6562958-B1.
XX
PD      13-MAY-2003.
XX
PF      04-JUN-1999; 99US-00328352.
XX
PR      09-JUN-1998; 98US-0088701P.
XX
PA      (GENO-) GENOME THERAPEUTICS CORP.
XX
PI      Breton G, Bush D;
XX
DR      WPI; 2003-576092/54.
DR      N-PSDB; ADA28954.
XX
PT      New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT      for diagnosing a bacterial disease, as components of antibacterial
PT      vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT      plants.
XX
PS      Example; SEQ ID NO 4367; 328pp; English.
XX
CC      The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC      The A. baumannii nucleic acids and polypeptides are useful as reagents
CC      for diagnosing a bacterial disease, as components of antibacterial
CC      vaccines, as targets for antibacterial drugs, to detect the presence of
CC      A. baumannii and other Acinetobacter species in a sample, in screening
CC      compounds for the ability to interfere with the A. baumannii life cycle
CC      or to inhibit A. baumannii infection, and as biocontrol agents for
CC      plants. The present sequence represents the amino acid sequence of an A.
CC      baumannii protein.
XX
SQ      Sequence 364 AA;
Query Match      0.8%; Score 8; DB 6; Length 364;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
QY      997 LSLITAIK 1004
      |||||
DB      164 LSLITAIK 171
      |||||
RESULT 63
ABP73447
```

ID ABP73447 standard; protein; 411 AA.  
XX  
AC ABP73447;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Candida albicans essential protein SEQ ID NO 7284.  
XX  
KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;  
KW signal transduction; DNA replication; cell division; growth;  
KW proliferation; Candida albicans; fungicide; antifungal.  
XX  
OS Candida albicans.  
XX  
PN WO200253728-A2.  
XX  
PD 11-JUL-2002.  
XX  
PF 26-DEC-2001; 2001WO-US049486.  
XX  
PR 29-DEC-2000; 2000US-0259128P.  
PR 20-FEB-2001; 2001US-00792024.  
PR 22-AUG-2001; 2001US-0314050P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KU;  
XX  
DR WPI; 2002-566694/60.  
DR N-PSDB; ABZ31997.  
XX  
PT Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele of  
PT a gene and placing other allele of the gene under conditional expression.  
XX  
PS Claim 44; SEQ ID NO 7284; 167pp + Sequence Listing; English.  
XX  
CC The invention relates to constructing (M1) a strain of diploid fungal  
CC cells in which both alleles of a gene are modified, comprising modifying  
CC one allele by insertion or replacement by a cassette having an  
CC expressible selectable marker and modifying other allele by  
CC recombination, of a promoter replacement fragment with a heterologous  
CC promoter, so that expression of the second allele is regulated by the  
CC promoter. (M1) is useful for constructing a strain of diploid fungal  
CC cells in which both alleles of a gene are modified. The diploid fungal  
CC cells having both alleles modified are useful for identifying a gene that  
CC is essential to the survival or growth of a fungus, a gene that  
CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
CC that contributes to the resistance and/or pathogenicity of a diploid fungus  
CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
CC and for identifying a therapeutic agent for treatment of a mammalian  
CC disease. (M1) is useful for identifying a compound which modulates the  
CC activity of a gene product, preferably enzymatic activity, carbon  
CC compound catabolism, biosynthetic, transporter, transcriptional,  
CC translational, signal transduction, DNA replication and cell division  
CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office  
XX  
SQ Sequence 411 AA;

Query Match 0.8%; Score 8; DB 5; Length 411;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 795 LEKELDLL 802  
| | | | | | | |  
DB 37 LEKELDLL 44

RESULT 64  
ADS28368  
ID ADS28368 standard; protein; 458 AA.  
XX  
AC ADS28368;  
XX  
DT 02-DEC-2004 (first entry)  
XX  
DE Bacterial polypeptide #17401.  
XX  
KW Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KW bacterial polypeptide.  
XX  
OS Bacteria.  
XX  
PN US2003233675-A1.  
XX  
PD 18-DEC-2003.  
XX  
PF 20-FEB-2003; 2003US-00369493.  
XX  
PR 21-FEB-2002; 2002US-0360039P.  
XX  
PA (CAOY/) CAO Y.  
PA (HINK/) HINKLE G J.  
PA (SLAT/) SLATER S C.  
PA (CHEN/) CHEN X.  
PA (GOLD/) GOLDMAN B S.  
XX  
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
XX  
DR WPI; 2004-061375/06.  
XX  
PT New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
PT microbial source, useful for producing plants with improved properties.  
XX  
PS Claim 1; SEQ ID NO 17401; 122pp; English.  
XX  
CC The invention relates to a recombinant DNA construct comprising a  
CC promoter functional in a plant cell, where the promoter is positioned to  
CC provide for expression of a polynucleotide encoding a polypeptide from a  
CC microbial source. The invention also relates to a transformed plant  
CC comprising the recombinant DNA construct and a method of producing a  
CC transformed plant having an improved property. The plant is a crop plant  
CC such as maize or soybean. The method of producing a transformed plant  
CC having an improved property comprises transforming a plant with the  
CC recombinant DNA construct and growing the transformed plant, where the  
CC polynucleotide or polypeptide is useful for improving plant properties.  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition. Improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polypeptide used in the  
CC scope of the invention. Note: The sequence data for this patent did not  
CC form part of the printed specification but was obtained in electronic  
CC format from USPTO at seqdata.uspto.gov/sequence.html.  
XX  
SQ Sequence 458 AA;

Query Match 0.8%; Score 8; DB 8; Length 458;







CC The construct is useful for transgenic plant seed which has in its genome  
CC the construct, that is functional in the plant to transcribe the oil-  
CC associated gene. The transgenic plant seed grows into a plant having  
CC enhanced seed oil as compared to wild type. The construct is useful for  
CC producing hybrid maize seed. The transgenic plant seed is useful for  
CC producing vegetable oil. The present sequence represents the amino acid  
CC sequence of an oil-associated gene related protein.

XX Sequence 488 AA;

Query Match 0.8%; Score 8; DB 8; Length 488;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811  
|||||  
DB 55 GLVEGKGP 62

RESULT 72  
ADC07838  
ID ADC07838 standard; protein; 496 AA.

AC ADC07838;

DT 18-DEC-2003 (first entry)

DE Rice protein sequence Seq ID104 related to grain filling.

XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;  
KW Carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;  
KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;  
KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;  
KW Gene; ds; plant.

XX Oryza sativa.

XX WO200300905-A2.

XX 03-JAN-2003.

XX 21-JUN-2002; 2002WO-IB002450.

XX 22-JUN-2001; 2001US-0300112P.

XX 26-SEP-2001; 2001US-0325277P.

XX 20-DEC-2001; 2001US-0342327P.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;

PI Glazebrook J, Katagiri F, Kreps J, Provart N, Rickes D;

PI N-PSDB; ADC07837.

DR WPI; 2003-229341/22.

XX Claim 1; SEQ ID NO 104; 130pp; English.

XX This invention, in the area of plant biotechnology, relates to novel  
CC polynucleotides comprising a nucleotide sequence encoding a protein which  
CC is involved in or associated with the synthesis, metabolism or  
CC degradation of carbohydrates in the plant grain and the expression of  
CC which is up-regulated during grain filling. The plant is selected from  
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,  
CC sugarbeet, wheat, and rice. The invention may be useful for the  
CC improvement of protein, oil, starch, fibre and moisture content of the  
CC cereal grains. In addition, carbohydrate levels may be modified to a more  
CC desirable level using the present invention. The present sequence is the  
CC amino acid sequence of a rice protein of the invention. Note: The

CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/publishedpct\_sequences.

XX Sequence 496 AA;

Query Match 0.8%; Score 8; DB 7; Length 496;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811  
|||||  
DB 55 GLVEGKGP 62

RESULT 73

ID ADJ50282 standard; protein; 500 AA.

XX ADJ50282;

XX 06-MAY-2004 (first entry)

DE Oil-associated gene related protein #1782.

XX oil-associated gene; transgenic; enhanced seed oil; vegetable oil.

XX Unidentified.

XX US2004025202-A1.

XX 05-FEB-2004.

XX 14-MAR-2003; 2003US-00389566.

XX 15-MAR-2002; 2002US-0365301P.

XX 26-JUN-2002; 2002US-0391786P.

XX (LAUR/) LAURIE C C.

XX (RAVA/) RAVANELLO M.

XX (SAVA/) SAVAGE T.

XX (LEDE/) LEDEUX J R.

XX (ROGE/) ROGERS J A.

XX Laurie CC, Ravanello M, Savage T, Ledoux JR, Rogers JA;

XX WPI; 2004-142683/14.

XX Novel recombinant DNA construct comprising a promoter functional in  
PI plants operably linked to an oil-associated gene for producing transgenic  
PI plant seed.

XX Example 3; SEQ ID NO 2386; 22pp; English.

XX The invention relates to a recombinant DNA construct comprising a  
CC promoter functional in plants operably linked to an oil-associated gene.  
CC The construct is useful for transgenic plant seed which has in its genome  
CC the construct, that is functional in the plant to transcribe the oil-  
CC associated gene. The transgenic plant seed grows into a plant having  
CC enhanced seed oil as compared to wild type. The construct is useful for  
CC producing hybrid maize seed. The transgenic plant seed is useful for  
CC producing vegetable oil. The present sequence represents the amino acid  
CC sequence of an oil-associated gene related protein.

XX Sequence 500 AA;

Query Match 0.8%; Score 8; DB 8; Length 500;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811  
|||||

XX	oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
KW	Unidentified.
XX	US2004025202-A1.
XX	OS
XX	05-FEB-2004.
XX	PD
XX	14-MAR-2003; 2003US-00389566.
XX	15-MAR-2002; 2002US-0365301P.
PR	26-JUN-2002; 2002US-0391786P.
PR	26-JUN-2002; 2002US-0392018P.
XX	(LAUR/) LAURIE C C.
PA	(RAVA/) RAVANELLO M.
PA	(SAVA/) SAVAGE T.
PA	(LEDE/) LEDEAUX J R.
PA	(ROGE/) ROGERS J A.
XX	Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
PI	WPI; 2004-142683/14.
DR	Novel recombinant DNA construct comprising a promoter functional in
XX	plants operably linked to an oil-associated gene for producing transgenic
PT	plant seed.
PT	Example 3; SEQ ID NO 860; 22pp; English.
XX	The invention relates to a recombinant DNA construct comprising a
CC	promoter functional in plants operably linked to an oil-associated gene.
CC	The construct is useful for transgenic plant seed which has in its genome
CC	the construct, that is functional in the plant to transcribe the oil-
CC	associated gene. The transgenic plant seed grows into a plant having
CC	enhanced seed oil as compared to wild type. The construct is useful for
CC	producing hybrid maize seed. The transgenic plant seed is useful for
CC	producing vegetable oil. The present sequence represents the amino acid
CC	sequence of an oil-associated gene related protein.
XX	Sequence 503 AA;
SQ	Query Match 0.8%; Score 8; DB 8; Length 503;
	Best Local Similarity 100.0%; Pred.No. 2.5e+02;
	Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Qy	804 GLVEGKGP 811
Db	55 GLVEGKGP 62
RESULT 76	
ADJ50184	
ID	ADJ50184 standard; protein; 503 AA.
XX	AC
XX	ADJ50184;
XX	06-MAY-2004 (first entry)
DT	Oil-associated gene related protein #1684.
XX	oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
DE	Unidentified.
XX	US2004025202-A1.
XX	OS
XX	05-FEB-2004.
PD	14-MAR-2003; 2003US-00389566.
XX	15-MAR-2002; 2002US-0365301P.
PR	





```

XX SQ Sequence 523 AA;
Query Match 0.8%; Score 8; DB 2; Length 523;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 51 GLVEGKGP 58

RESULT 80
ADJ48855
ID ADJ48855 standard; protein; 527 AA.
XX
AC ADJ48855;
XX
DT 06-MAY-2004 (first entry)
XX
DE Oil-associated gene related protein #355.
XX
KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX
OS Unidentified.
XX
PN US2004025202-A1.
XX
PD 05-FEB-2004.
XX
PF 14-MAR-2003; 2003US-00389566.
XX
PR 15-MAR-2002; 2002US-0365301P.
XX
PR 26-JUN-2002; 2002US-0391786P.
XX
PR 26-JUN-2002; 2002US-0392018P.
XX
(LAUR/) LAURIE C C.
PA (RAVA/) RAVANELLO M.
PA (SAVA/) SAVAGE T.
PA (LEDE/) LEDEAUX J R.
PA (ROGE/) ROGERS J A.
XX
PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
XX
WPI; 2004-142683/14.
XX
PT Novel recombinant DNA construct comprising a promoter functional in
PT plants operably linked to an oil-associated gene for producing transgenic
PT plant seed.
XX
PS Example 3; SEQ ID NO 859; 22pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in plants operably linked to an oil-associated gene.
CC The construct is useful for transgenic plant seed which has in its genome
CC the construct, that is functional in the plant to transcribe the oil-
CC associated gene. The transgenic plant seed grows into a plant having
CC enhanced seed oil as compared to wild type. The construct is useful for
CC producing hybrid maize seed. The transgenic plant seed is useful for
CC producing vegetable oil. The present sequence represents the amino acid
CC sequence of an oil-associated gene related protein.
XX
SQ Sequence 527 AA;
Query Match 0.8%; Score 8; DB 8; Length 527;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
Db 57 GLVEGKGP 64

RESULT 81
AAG31382
ID AAG31382 standard; protein; 528 AA.
XX
AC AAG31382;

```



```

PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      0.8%; Score 8; DB 3; Length 528;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 LKTVALT 12
DB      509 LKTVALT 516

RESULT 82
AAAG31381
ID AAG31381 standard; protein; 529 AA.
XX
AC AAG31381;
XX
DT
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37676.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
```

termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-00301439.

25-FEB-1999; 99US-0121825P.

05-MAR-1999; 99US-0123180P.

09-MAR-1999; 99US-0123548P.

23-MAR-1999; 99US-0125789P.

25-MAR-1999; 99US-0126264P.

29-MAR-1999; 99US-0126785P.

01-APR-1999; 99US-0127462P.

06-APR-1999; 99US-0128234P.

08-APR-1999; 99US-0128714P.

16-APR-1999; 99US-0129845P.

19-APR-1999; 99US-0130077P.

21-APR-1999; 99US-0130449P.

23-APR-1999; 99US-0130510P.

23-APR-1999; 99US-0130891P.

28-APR-1999; 99US-0131449P.

30-APR-1999; 99US-0132048P.

04-MAY-1999; 99US-0132407P.

05-MAY-1999; 99US-0132484P.

06-MAY-1999; 99US-0132485P.

06-MAY-1999; 99US-0132486P.

07-MAY-1999; 99US-0132863P.

11-MAY-1999; 99US-0132867P.

14-MAY-1999; 99US-0134218P.

14-MAY-1999; 99US-0134219P.

14-MAY-1999; 99US-0134221P.

18-MAY-1999; 99US-0134370P.

19-MAY-1999; 99US-0134768P.

20-MAY-1999; 99US-0134941P.

21-MAY-1999; 99US-0135124P.

21-MAY-1999; 99US-0135353P.

24-MAY-1999; 99US-0135629P.

25-MAY-1999; 99US-0136021P.

27-MAY-1999; 99US-0136392P.

28-MAY-1999; 99US-0136782P.

01-JUN-1999; 99US-0137222P.

03-JUN-1999; 99US-0137528P.

04-JUN-1999; 99US-0137502P.

07-JUN-1999; 99US-0137724P.

08-JUN-1999; 99US-0138094P.

10-JUN-1999; 99US-0138540P.

10-JUN-1999; 99US-0138847P.

14-JUN-1999; 99US-0139119P.

16-JUN-1999; 99US-0139452P.

16-JUN-1999; 99US-0139453P.

17-JUN-1999; 99US-0139453P.

17-JUN-1999; 99US-0139492P.

18-JUN-1999; 99US-0139454P.

18-JUN-1999; 99US-0139455P.

18-JUN-1999; 99US-0139456P.

18-JUN-1999; 99US-0139457P.

18-JUN-1999; 99US-0139458P.

18-JUN-1999; 99US-0139459P.

18-JUN-1999; 99US-0139460P.

18-JUN-1999; 99US-0139461P.

18-JUN-1999; 99US-0139462P.

18-JUN-1999; 99US-0139463P.

18-JUN-1999; 99US-0139750P.

21-JUN-1999; 99US-0139763P.

21-JUN-1999; 99US-0139817P.

22-JUN-1999; 99US-0139899P.

23-JUN-1999; 99US-0140353P.

23-JUN-1999; 99US-0140354P.

24-JUN-1999; 99US-0140695P.

```
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.

PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157855P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 0.8%; Score 8; DB 3; Length 529;
Best Local Similarity 100.0%; Pred.No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTL 12
Db 510 LKTVALTL 517

RESULT 83
AAR67005
ID AAR67005 standard; protein; 530 AA.
XX
AC AAR67005;
XX
DT 17-AUG-1995 (first entry)
XX
DE Barley beta-amylase.
XX
KW Barley beta-amylase; recombinant production.
XX
OS Hordeum vulgare.
XX
PN JP06303983-A.
XX
PD 01-NOV-1994.
XX
PF 18-FEB-1994; 94JP-00058119.
```

```
XX 26-FEB-1993; 93JP-00038870.
XX (SAPB ) SAPPORO BREWERIES.
XX WPI; 1995-018276/03.
XX DR N-PSDB; AAQ74813.
XX
XX Barley beta-amylase gene - for the efficient production of beta-amylase.
XX
XX Claim 1; Page 12-13; 19pp; Japanese.
XX
XX AAQ74813 encodes AAR67005 barley beta-amylase. AAQ74813 can be used in
XX the construction of a vector, for the recombinant expression of barley
XX beta-amylase
XX
XX Sequence 530 AA;
SQ
Query Match 0.8%; Score 8; DB 2; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62
|||||
55 GLVEGKGP 62

RESULT 84
AAR97613
ID AAR97613 standard; protein; 531 AA.
XX AC
XX AAR97613;
XX
XX 20-AUG-1996 (first entry)
XX
XX Beta-amylase sevenfold mutant.
XX
XX Beta-amylase; maltose; thermostable enzyme; protein engineering;
XX enzyme engineering; barley; Hordeum vulgare.
XX
XX Synthetic.
XX
XX EP713916-A2.
XX
XX 29-MAY-1996.
XX
XX 27-SEP-1995; 95EP-00115255.
XX
XX 28-SEP-1994; 94JP-00233086.
XX
XX (SAPB ) SAPPORO BREWERIES.
XX
XX Yoshigi N, Maeba H, Okada Y;
XX
XX WPI; 1996-253043/26.
XX DR N-PSDB; AAT29192.
XX
XX Recombinant beta-amylase used in the prodn. of maltose - has superior
XX thermostability compared to wild type enzymes.
XX
XX Claim 2; Page 18-19; 29pp; English.
XX
XX A 7-fold mutant beta-amylase (AAR97613) has a thermostability improved by
XX 11.6 degC compared with the native (barley) enzyme and 5.8 degC compared
XX with soybean beta-amylase. It contains substitutions of Leu for Met-181,
XX Ala for Ser-291, Val for Ile-293, Pro for Ser-347, Pro for Ser-346, Asp
XX for Gln-348 and Ser for Ala-372. The mutant beta-amylase can be expressed
XX in transformed Escherichia coli JM109 cells carrying a vector (see also
XX AAT29193) that incorporates the mutated coding sequence (AAT29192). It is
XX useful for industrial prodn. of maltose
XX
XX Sequence 531 AA;
SQ
Query Match 0.8%; Score 8; DB 2; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62
|||||
55 GLVEGKGP 62

RESULT 84
AAR97613
ID AAR97613 standard; protein; 531 AA.
XX AC
XX AAR97613;
XX
XX 20-AUG-1996 (first entry)
XX
XX Beta-amylase sevenfold mutant.
XX
XX Beta-amylase; maltose; thermostable enzyme; protein engineering;
XX enzyme engineering; barley; Hordeum vulgare.
XX
XX Synthetic.
XX
XX EP713916-A2.
XX
XX 29-MAY-1996.
XX
XX 27-SEP-1995; 95EP-00115255.
XX
XX 28-SEP-1994; 94JP-00233086.
XX
XX (SAPB ) SAPPORO BREWERIES.
XX
XX Yoshigi N, Maeba H, Okada Y;
XX
XX WPI; 1996-253043/26.
XX DR N-PSDB; AAT29192.
XX
XX Recombinant beta-amylase used in the prodn. of maltose - has superior
XX thermostability compared to wild type enzymes.
XX
XX Claim 2; Page 18-19; 29pp; English.
XX
XX A 7-fold mutant beta-amylase (AAR97613) has a thermostability improved by
XX 11.6 degC compared with the native (barley) enzyme and 5.8 degC compared
XX with soybean beta-amylase. It contains substitutions of Leu for Met-181,
XX Ala for Ser-291, Val for Ile-293, Pro for Ser-347, Pro for Ser-346, Asp
XX for Gln-348 and Ser for Ala-372. The mutant beta-amylase can be expressed
XX in transformed Escherichia coli JM109 cells carrying a vector (see also
XX AAT29193) that incorporates the mutated coding sequence (AAT29192). It is
XX useful for industrial prodn. of maltose
XX
XX Sequence 531 AA;
SQ
Query Match 0.8%; Score 8; DB 2; Length 531;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 804 GLVEGKGP 811
DB 51 GLVEGKGP 58
|||||
51 GLVEGKGP 58

RESULT 85
ADJ49017
ID ADJ49017 standard; protein; 533 AA.
XX AC
XX ADJ49017;
XX
XX 06-MAY-2004 (first entry)
XX
XX Oil-associated gene related protein #517.
XX
XX oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX
XX Unidentified.
XX
XX US2004025202-A1.
XX
XX 05-FEB-2004.
XX
XX 14-MAR-2003; 2003US-00389566.
XX
XX 15-MAR-2002; 2002US-0365301P.
XX PR 26-JUN-2002; 2002US-0391786P.
XX PR 26-JUN-2002; 2002US-0392018P.
XX
XX (LAUR/) LAURIE C C.
XX PA (RAVA/) RAVANELLO M.
XX PA (SAVA/) SAVAGE T.
XX PA (LEDE/) LEDEAUX J R.
XX PA (ROGE/) ROGERS J A.
XX
XX Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
XX
XX WPI; 2004-142683/14.
XX
XX Novel recombinant DNA construct comprising a promoter functional in
XX plants operably linked to an oil-associated gene for producing transgenic
XX plant seed.
XX
XX Example 3; SEQ ID NO 1021; 22pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in plants operably linked to an oil-associated gene.
XX The construct is useful for transgenic plant seed which has in its genome
XX the construct, that is functional in the plant to transcribe the oil-
XX associated gene. The transgenic plant seed grows into a plant having
XX enhanced seed oil as compared to wild type. The construct is useful for
XX producing hybrid maize seed. The transgenic plant seed is useful for
XX producing vegetable oil. The present sequence represents the amino acid
XX sequence of an oil-associated gene related protein.
XX
XX Sequence 533 AA;
SQ
Query Match 0.8%; Score 8; DB 8; Length 533;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 804 GLVEGKGP 811
DB 53 GLVEGKGP 60
|||||
53 GLVEGKGP 60

RESULT 86
ADJ49016
ID ADJ49016 standard; protein; 533 AA.
```

```

XX AC ADJ49016;
XX DT 06-MAY-2004 (first entry)
XX DE Oil-associated gene related protein #516.
XX KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX OS Unidentified.
XX PN US2004025202-A1.
XX PD 05-FEB-2004.
XX PF 14-MAR-2003; 2003US-00389566.
XX PR 15-MAR-2002; 2002US-0365301P.
XX PR 26-JUN-2002; 2002US-0391786P.
XX PR 26-JUN-2002; 2002US-0394018P.
XX PA (LAUR/) LAURIE C C.
XX PA (RAVA/) RAVANELLO M.
XX PA (SAVA/) SAVAGE T.
XX PA (LEDE/) LEDEUX J R.
XX PA (ROGE/) ROGERS J A.
XX FI Laurie CC, Ravanello M, Savage T, Ledoux JR, Rogers JA;
XX WI; 2004-142683/14.
XX DR
XX PT Novel recombinant DNA construct comprising a promoter functional in
XX plants operably linked to an oil-associated gene for producing transgenic
XX plant seed.
XX PS Example 3; SEQ ID NO 1020; 22pp; English.
XX CC The invention relates to a recombinant DNA construct comprising a
XX promoter functional in plants operably linked to an oil-associated gene.
XX CC The construct is useful for transgenic plant seed which has in its genome
XX the construct, that is functional in the plant to transcribe the oil-
XX associated gene. The transgenic plant seed grows into a plant having
XX enhanced seed oil as compared to wild type. The construct is useful for
XX producing hybrid maize seed. The transgenic plant seed is useful for
XX producing vegetable oil. The present sequence represents the amino acid
XX sequence of an oil-associated gene related protein.
XX SQ Sequence 533 AA;

Query Match 0.8%; Score 8; DB 8; Length 533;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
DB 53 GLVEGKGP 60
|||||

RESULT 87
AAW04261
ID AAW04261 standard; protein; 535 AA.
XX AC AAW04261;
XX DT 16-JUN-1997 (first entry)
XX DE Beta-amylase protein.
XX KW Beta-amylase; barley; structural protein; thermally stable enzyme.
XX OS Hordeum vulgare.
XX PN WO9630525-A1.

```

```

XX PD 03-OCT-1996.
XX PF 27-MAR-1996; 96WO-JP000799.
XX PR 27-MAR-1995; 95JP-00092004.
XX PA (SAPB ) SAPPORO BREWERIES.
XX PI Yoshigi N, Okada Y;
XX WI; 1996-455370/45.
XX DR N-PSDB; AAT33961.
XX PT Barley beta-amylase structural gene - for prodn. of recombinant beta-
XX amylase, and plants with improved characteristics, useful in beverage and
XX food industries.
XX PS Claim 1; Page 8-11; 30pp; Japanese.
XX CC This sequence represents the barley beta-amylase structural protein. The
XX sequence represented by AAW15746 represents a mutated version of this
XX sequence. The DNA encoding this sequence was isolated from barley seed
XX genomic DNA using the primers represented by AAT33963 and AAT33964. The
XX gene encoding this sequence is used for the production of recombinant
XX beta-amylase with good thermal stability. The gene is also used to create
XX transformants of barley and other plants with improved characteristics.
XX CC The transformants can then be used in the alcoholic beverage (e.g. beer
XX and distilled spirits), food and enzyme industries
XX SQ Sequence 535 AA;

Query Match 0.8%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811
DB 55 GLVEGKGP 62
|||||

RESULT 88
AAW15746
ID AAW15746 standard; protein; 535 AA.
XX AC AAW15746;
XX DT 16-JUN-1997 (first entry)
XX DE Mutated beta-amylase protein.
XX KW Beta-amylase; barley; structural protein; thermally stable enzyme.
XX OS Hordeum vulgare.
XX PI Key Location/Qualifiers
XX FT Misc-difference 185 /label= M185L
XX FT Misc-difference 295 /label= S295A
XX FT Misc-difference 297 /label= I297V
XX FT Misc-difference 350 /label= S350P
XX FT Misc-difference 351 /label= S351P
XX FT Misc-difference 352 /label= Q352D
XX FT Misc-difference 376 /label= A376S
XX PN WO9630525-A1.

```



PD 03-OCT-1996.  
 XX  
 PF 27-MAR-1996; 96WO-JP000799.  
 XX  
 PR 27-MAR-1995; 95JP-00092004.  
 XX  
 PA (SAPB ) SAPPORO BREWERIES.  
 XX  
 PI Yoshigi N, Okada Y;  
 XX  
 DR WPI; 1996-455370/45.  
 DR N-PSDB; AA133962.  
 XX  
 PT Barley beta-amylase structural gene - for prodn. of recombinant beta-  
 PT amylase, and plants with improved characteristics, useful in beverage and  
 PT food industries.  
 XX  
 PS Claim 2; Page; 30pp; Japanese.  
 XX  
 CC This sequence represents a mutated version of the barley beta-amylase  
 CC structural protein. The sequence represented by AA15745 represents the  
 CC wild type version of this sequence. The DNA encoding the wild type  
 CC sequence was isolated from barley seed genomic DNA using the primers  
 CC represented by AA133963 and AA133964. The gene encoding this sequence is  
 CC used for the production of recombinant beta-amylase with good thermal  
 CC stability. The gene is also used to create transformants of barley and  
 CC other plants with improved characteristics. The transformants can then be  
 CC used in the alcoholic beverage (e.g. beer and distilled spirits), food  
 CC and enzyme industries  
 XX  
 SQ Sequence 535 AA;  
 Query Match 0.8%; Score 8; DB 2; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 804 GLVEGKGP 811  
 Db |||||  
 55 GLVEGKGP 62  
 RESULT 89  
 ADJ49120  
 ID ADJ49120 standard; protein; 535 AA.  
 AC  
 AC ADJ49120;  
 XX  
 XX 06-MAY-2004 (first entry)  
 DT  
 DE Oil-associated gene related protein #620.  
 XX  
 XX oil-associated gene; transgenic; enhanced seed oil; vegetable oil.  
 XX Unidentified.  
 OS  
 XX US2004025202-A1.  
 PN  
 PD 05-FEB-2004.  
 XX  
 PF 14-MAR-2003; 2003US-00389566.  
 XX  
 PR 15-MAR-2002; 2002US-0365301P.  
 PR 26-JUN-2002; 2002US-0391786P.  
 PR 26-JUN-2002; 2002US-0392018P.  
 XX  
 PA (LAUR/) LAURIE C C.  
 PA (RAVA/) RAVANELLO M.  
 PA (SAVA/) SAVAGE T.  
 PA (LEDE/) LEDEAUX J R.  
 PA (ROGE/) ROGERS J A.  
 XX  
 PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;  
 XX  
 XX WPI; 2004-142683/14.  
 DR  
 XX Novel recombinant DNA construct comprising a promoter functional in  
 PT plants operably linked to an oil-associated gene for producing transgenic  
 PT plant seed.  
 XX  
 PS Example 3; SEQ ID NO 2397; 22pp; English.  
 XX  
 CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in plants operably linked to an oil-associated gene.  
 CC The construct is useful for transgenic plant seed which has in its genome  
 CC the construct, that is functional in the plant to transcribe the oil-  
 CC associated gene. The transgenic plant seed grows into a plant having  
 CC enhanced seed oil as compared to wild type. The construct is useful for  
 CC producing hybrid maize seed. The transgenic plant seed is useful for  
 CC producing vegetable oil. The present sequence represents the amino acid  
 CC sequence of an oil-associated gene related protein.

DR WPI; 2004-142683/14.  
 XX  
 PT Novel recombinant DNA construct comprising a promoter functional in  
 PT plants operably linked to an oil-associated gene for producing transgenic  
 PT plant seed.  
 XX  
 PS Example 3; SEQ ID NO 1124; 22pp; English.  
 XX  
 CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in plants operably linked to an oil-associated gene.  
 CC The construct is useful for transgenic plant seed which has in its genome  
 CC the construct, that is functional in the plant to transcribe the oil-  
 CC associated gene. The transgenic plant seed grows into a plant having  
 CC enhanced seed oil as compared to wild type. The construct is useful for  
 CC producing hybrid maize seed. The transgenic plant seed is useful for  
 CC producing vegetable oil. The present sequence represents the amino acid  
 CC sequence of an oil-associated gene related protein.  
 XX  
 SQ Sequence 535 AA;  
 Query Match 0.8%; Score 8; DB 8; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 804 GLVEGKGP 811  
 Db |||||  
 55 GLVEGKGP 62  
 RESULT 90  
 ADJ50393  
 ID ADJ50393 standard; protein; 535 AA.  
 XX  
 AC ADJ50393;  
 XX  
 XX 06-MAY-2004 (first entry)  
 DT  
 DE Oil-associated gene related protein #1893.  
 XX  
 XX oil-associated gene; transgenic; enhanced seed oil; vegetable oil.  
 XX Unidentified.  
 OS  
 XX US2004025202-A1.  
 PN  
 PD 05-FEB-2004.  
 XX  
 PF 14-MAR-2003; 2003US-00389566.  
 XX  
 PR 15-MAR-2002; 2002US-0365301P.  
 PR 26-JUN-2002; 2002US-0391786P.  
 PR 26-JUN-2002; 2002US-0392018P.  
 XX  
 PA (LAUR/) LAURIE C C.  
 PA (RAVA/) RAVANELLO M.  
 PA (SAVA/) SAVAGE T.  
 PA (LEDE/) LEDEAUX J R.  
 PA (ROGE/) ROGERS J A.  
 XX  
 PI Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;  
 XX  
 XX WPI; 2004-142683/14.  
 DR  
 XX Novel recombinant DNA construct comprising a promoter functional in  
 PT plants operably linked to an oil-associated gene for producing transgenic  
 PT plant seed.  
 XX  
 PS Example 3; SEQ ID NO 2397; 22pp; English.  
 XX  
 CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in plants operably linked to an oil-associated gene.  
 CC The construct is useful for transgenic plant seed which has in its genome  
 CC the construct, that is functional in the plant to transcribe the oil-

CC associated gene. The transgenic plant seed grows into a plant having  
 CC enhanced seed oil as compared to wild type. The construct is useful for  
 CC producing hybrid maize seed. The transgenic plant seed is useful for  
 CC producing vegetable oil. The present sequence represents the amino acid  
 CC sequence of an oil-associated gene related protein.  
 XX  
 XX

SQ Sequence 535 AA;

Query Match 0.8%; Score 8; DB 8; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811  
 |||||  
 Db 55 GLVEGKGP 62

RESULT 91  
 ADJ50119  
 ID ADJ50119 standard; protein; 535 AA.  
 XX  
 AC ADJ50119;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Oil-associated gene related protein #1619.  
 XX  
 KW oil-associated gene; transgenic; enhanced seed oil; vegetable oil.  
 XX  
 OS Unidentified.  
 XX  
 PN US2004025202-A1.  
 XX  
 PD 05-FEB-2004.  
 XX  
 PF 14-MAR-2003; 2003US-00389566.  
 XX  
 PR 15-MAR-2002; 2002US-0365301P.  
 PR 26-JUN-2002; 2002US-0391786P.  
 PR 26-JUN-2002; 2002US-0392018P.  
 XX

(LAUR/) LAURIE C C.  
 (RAVA/) RAVANELLO M.  
 (SAVA/) SAVAGE T.  
 (LEDE/) LEDEAUX J R.  
 (ROGE/) ROGERS J A.  
 XX

Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;  
 WPI; 2004-142683/14.  
 XX  
 DR Novel recombinant DNA construct comprising a promoter functional in  
 PT plants operably linked to an oil-associated gene for producing transgenic  
 PT plant seed.  
 XX  
 PS Example 3; SEQ ID NO 2123; 22pp; English.  
 XX  
 CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in plants operably linked to an oil-associated gene.  
 CC The construct is useful for transgenic plant seed which has in its genome  
 CC the construct, that is functional in the plant to transcribe the oil-  
 CC associated gene. The transgenic plant seed grows into a plant having  
 CC enhanced seed oil as compared to wild type. The construct is useful for  
 CC producing hybrid maize seed. The transgenic plant seed is useful for  
 CC producing vegetable oil. The present sequence represents the amino acid  
 CC sequence of an oil-associated gene related protein.  
 XX

SQ Sequence 535 AA;

Query Match 0.8%; Score 8; DB 8; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811  
 |||||  
 Db 55 GLVEGKGP 62

RESULT 92

ADJ49037  
 ID ADJ49037 standard; protein; 535 AA.  
 XX  
 AC ADJ49037;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX

DE Oil-associated gene related protein #537.

XX oil-associated gene; transgenic; enhanced seed oil; vegetable oil.

XX Unidentified.

OS US2004025202-A1.

XX PD 05-FEB-2004.

XX PF 14-MAR-2003; 2003US-00389566.

XX PR 15-MAR-2002; 2002US-0365301P.

PR 26-JUN-2002; 2002US-0391786P.

PR 26-JUN-2002; 2002US-0392018P.

XX (LAUR/) LAURIE C C.

PA (RAVA/) RAVANELLO M.

PA (SAVA/) SAVAGE T.

PA (LEDE/) LEDEAUX J R.

PA (ROGE/) ROGERS J A.

XX Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;

XX WPI; 2004-142683/14.

XX Novel recombinant DNA construct comprising a promoter functional in

PT plants operably linked to an oil-associated gene for producing transgenic

PT plant seed.

XX Example 3; SEQ ID NO 1041; 22pp; English.

XX The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in plants operably linked to an oil-associated gene.  
 CC The construct is useful for transgenic plant seed which has in its genome  
 CC the construct, that is functional in the plant to transcribe the oil-  
 CC associated gene. The transgenic plant seed grows into a plant having  
 CC enhanced seed oil as compared to wild type. The construct is useful for  
 CC producing hybrid maize seed. The transgenic plant seed is useful for  
 CC producing vegetable oil. The present sequence represents the amino acid  
 CC sequence of an oil-associated gene related protein.  
 XX

SQ Sequence 535 AA;

Query Match 0.8%; Score 8; DB 8; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 804 GLVEGKGP 811  
 |||||  
 Db 55 GLVEGKGP 62

RESULT 93

ADJ50182

ID ADJ50182 standard; protein; 535 AA.

XX

AC ADJ50182;

XX DT 06-MAY-2004 (first entry)

XX	DE	Oil-associated gene related protein #1682.
XX	KW	oil-associated gene; transgenic; enhanced seed oil; vegetable oil.
XX	OS	Unidentified.
XX	PN	US2004025202-A1.
XX	PD	05-FEB-2004.
XX	PP	14-MAR-2003; 2003US-00389566.
XX	PR	15-MAR-2002; 2002US-0365301P.
XX	PR	26-JUN-2002; 2002US-0391786P.
XX	PR	26-JUN-2002; 2002US-0392018P.
XX	PA	(LAUR/) LAURIE C C.
XX	PA	(RAVA/) RAVANELLO M.
XX	PA	(SAVA/) SAVAGE T.
XX	PA	(LEDE/) LEDEAUX J R.
XX	PA	(ROGE/) ROGERS J A.
XX	PI	Laurie CC, Ravanello M, Savage T, Ledeaux JR, Rogers JA;
XX	DR	WPI; 2004-142683/14.
XX	PT	Novel recombinant DNA construct comprising a promoter functional in
XX	PT	plants operably linked to an oil-associated gene for producing transgenic
XX	PT	plant seed.
XX	PS	Example 3; SEQ ID NO 2186; 22pp; English.
XX	CC	The invention relates to a recombinant DNA construct comprising a
XX	CC	promoter functional in plants operably linked to an oil-associated gene.
XX	CC	The construct is useful for transgenic plant seed which has in its genome
XX	CC	the construct, that is functional in the plant to transcribe the oil-
XX	CC	associated gene. The transgenic plant seed grows into a plant having
XX	CC	enhanced seed oil as compared to wild type. The construct is useful for
XX	CC	producing hybrid maize seed. The transgenic plant seed is useful for
XX	CC	producing vegetable oil. The present sequence represents the amino acid
XX	CC	sequence of an oil-associated gene related protein.
XX	SQ	Sequence 535 AA;
Query Match		
Best Local Similarity 0.8%; Score 8; DB 8; Length 535;		
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0		
Qy	804	GLVEGKGP 811
Dd	55	GLVEGKGP 62
RESULT 94		
AAB07724		
ID	AAB07724	standard; protein; 555 AA.
AC	AAB07724;	
XX	07-NOV-2000	(first entry)
XX	An Arabidopsis aintegumenta (ANT) polypeptide.	
XX	Aintegumenta; ANT; cell proliferation; growth; organ mass; fertility;	
XX	asexual reproduction; plant; male sterile plant; female sterile plant;	
XX	early flowering.	
XX	Arabidopsis sp.	
XX	Key	Location/Qualifiers
XX	Domain	281..357
XX	FT	/note= "AP2 domain"

PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131443P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135623P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144326P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145132P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151085P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.

PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161922P.  
 PR 28-OCT-1999; 99US-0161993P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 0.8%; Score 8; DB 3; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTTL 12  
 |||||  
 Db 536 LKTVALTTL 543

RESULT 96  
 ABR40832  
 ID ABR40832 standard; protein; 555 AA.  
 AC ABR40832;

16-MAY-2003 (first entry)  
 DE Arabidopsis thaliana oil trait related protein sequence SEQ ID NO:424.  
 XX Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;  
 KW receptor-like protein kinase; mitogen activated protein kinase; oil;  
 KW Lip15-like transcription factor caleosin; ATP citrate lyase; SNF1;  
 KW CKC-like transcription factor; antisense inhibition; co-suppression;  
 KW transgenic plant.

OS Arabidopsis thaliana.

XX WO2003002751-A2.

XX 09-JAN-2003.

XX 27-JUN-2002; 2002WO-US020152.

XX 29-JUN-2001; 2001US-0301913P.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PA (PION-) PIONEER HI-BRED INT INC.

XX Allen SM, Allen WB, Cahoon RE, Epelbaum S, Famodu OO, Harvell LT;  
 PI Jones TJ, Kinney AJ, Klein TM, Li C, Oliveira IC, Sakai H, Shen B;  
 PI Tarczynski MC;  
 XX WPI; 2003-201509/19.

PT Novel nucleotide fragment encoding polypeptides having receptor-like  
 PT protein kinase activity, caleosin-like activity, useful for altering oil  
 PT phenotypes in plants such as sunflower, coconut, soybean, wheat and rice.  
 XX Claim 12; Page 463-465; 542pp; English.  
 XX The present invention describes an isolated nucleotide fragment (I)  
 CC comprising a nucleic acid sequence (NS) chosen from a NS encoding a  
 CC polypeptide (PP) having receptor-like protein kinase activity, mitogen  
 CC activated protein (MAP)-kinase activity, Lip15-like transcription factor  
 CC activity, caleosin-like activity, ATP citrate lyase activity, SNF1-like  
 CC activity and CKC-like transcription factor activity. Also described: (1)  
 CC complement (II) of (I); (2) a chimeric construct (III) comprising (I) or  
 CC (II), operably linked to a regulatory sequence; (3) a plant (IV)  
 CC comprising (III) in its genome; (4) seeds (V) obtained from (IV); and (5)  
 CC oil obtained from (V). (I) or its part can be used in antisense  
 CC inhibition or co-suppression in a transformed plant. (III) is useful for  
 CC altering the oil phenotype in a plant such as corn, soybean, wheat, rice,  
 CC canola, Brassica, sorghum, sunflower or coconut. (III) is also useful for  
 CC creating transgenic plants having altered lipid profiles. (I) can also be  
 CC used as a hybridisation probe. ACC00626 to ACC00868 and ABR40591 to  
 CC ABR40879 represent sequences used in the exemplification of the present  
 CC invention  
 XX Sequence 555 AA;

Query Match 0.8%; Score 8; DB 6; Length 555;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTTL 12  
 |||||  
 Db 536 LKTVALTTL 543

RESULT 97  
 ABR40830  
 ID ABR40830 standard; protein; 555 AA.  
 XX ABR40830;

XX 16-MAY-2003 (first entry)

XX Arabidopsis thaliana oil trait related protein sequence SEQ ID NO:422.  
 KW Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;  
 KW receptor-like protein kinase; mitogen activated protein kinase; oil;  
 KW Lip15-like transcription factor caleosin; ATP citrate lyase; SNF1;  
 KW CKC-like transcription factor; antisense inhibition; co-suppression;  
 KW transgenic plant.

OS Arabidopsis thaliana.

XX WO2003002751-A2.

XX 09-JAN-2003.

XX 27-JUN-2002; 2002WO-US020152.

XX 29-JUN-2001; 2001US-0301913P.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PA (PION-) PIONEER HI-BRED INT INC.

XX Allen SM, Allen WB, Cahoon RE, Epelbaum S, Famodu OO, Harvell LT;  
 PI Jones TJ, Kinney AJ, Klein TM, Li C, Oliveira IC, Sakai H, Shen B;  
 PI Tarczynski MC;  
 XX WPI; 2003-201509/19.

PT Novel nucleotide fragment encoding polypeptides having receptor-like  
 PT protein kinase activity, caleosin-like activity, useful for altering oil  
 PT phenotypes in plants such as sunflower, coconut, soybean, wheat and rice.

XX Claim 12; Page 460-461; 542pp; English.

PS The present invention describes an isolated nucleotide fragment (I)

XX comprising a nucleic acid sequence (NS) chosen from a NS encoding a

CC polypeptide (PP) having receptor-like protein kinase activity, mitogen

CC activated protein (MAP)-kinase activity, LIP5-like transcription factor

CC activity, calceosin-like activity, ATP citrate lyase activity, SNF1-like

CC activity, and CKC-like transcription factor activity. Also described: (1)

CC complement (II) of (I); (2) a chimeric construct (III) comprising (I) or

CC (II), operably linked to a regulatory sequence; (3) a plant (IV)

CC comprising (III) in its genome; (4) seeds (V) obtained from (IV); and (5)

CC oil obtained from (V). (I) or its part can be used in antisense

CC inhibition or co-suppression in a transformed plant. (III) is useful for

CC altering the oil phenotype in a plant such as corn, soybean, wheat, rice,

CC canola, Brassica, sorghum, sunflower or coconut. (III) is also useful for

CC creating transgenic plants having altered lipid profiles. (I) can also be

CC used as a hybridisation probe. ACC00626 to ACC00868 and ABR40591 to

CC ABR40879 represent sequences used in the exemplification of the present

CC invention

XX Sequence 555 AA;

SQ

Query Match 0.8%; Score 8; DB 6; Length 555;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LKTVALTL 12

Db 536 LKTVALTL 543

|||||

RESULT 98

ID ADG25136 standard; protein; 555 AA.

AC ADG25136;

XX

XX 26-FEB-2004 (first entry)

DT

DE Thalecress Aintegumenta, ANT.

XX

KW Thalecress; Aintegumenta; ANT; AP2 domain; plant; seed mass;

KW asexual reproduction; reduced fertility.

XX

XX Arabidopsis thaliana.

OS

XX US6639128-B1

PN

XX 28-OCT-2003.

PD

XX 07-JAN-2000; 2000US-00479855.

PF

XX 08-JAN-1999; 99US-00227421.

PR

XX (NASC-) NAT SCI FOUND.

PA

XX Fischer RL, Mizukami Y;

PI

XX WPI; 2003-842795/78.

DR

XX N-PSDB; ADG25135.

XX

XX New nucleic acid, useful in conferring desired traits on plants, such as

PT increased seed mass, asexual reproduction or reduced fertility.

PT

XX Disclosure; SEQ ID NO 2; 25pp; English.

PS

XX The invention relates to an isolated nucleic acid comprising a

CC polynucleotide encoding a polypeptide having at least 90% identity with

CC Canola Aintegumenta, ANT, (appearing as ADG25139) as determined using a

CC BLAST algorithm and comprising two AP2 domains. Also included are an

CC expression cassette comprising a plant promoter operably linked to a

CC heterologous polynucleotide encoding the polypeptide, a vector comprising

CC the expression cassette, a plant comprising the plant promoter operably

CC linked to the heterologous polynucleotide, a method of asexually

CC reproducing a plant and a method of reducing fertility in a plant. The

CC nucleic acid is useful in conferring desired traits on plants, such as

CC increased seed mass, asexual reproduction or reduced fertility. The

CC present sequence represents Thalecress Aintegumenta, ANT.

XX Sequence 555 AA;

SQ

Query Match 0.8%; Score 8; DB 7; Length 555;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LKTVALTL 12

Db 536 LKTVALTL 543

|||||

RESULT 99

ID ADG88451 standard; protein; 555 AA.

AC ADG88451;

XX

XX 11-MAR-2004 (first entry)

DT

XX Arabidopsis thaliana aintegumenta (ANT) mutant protein, ANTMRII.

DE

KW Plant; genetic engineering; cell proliferation; aintegumenta; ANT;

KW transgenic; transgenic plant; organ mass alteration; fertility;

KW asexual reproduction; mutant; muten.

XX

OS Synthetic.

OS Arabidopsis thaliana.

XX

XX US2003159180-A1.

PN

XX 21-AUG-2003.

PD

XX 28-JAN-2002; 2002US-00059911.

PF

XX 28-JAN-2002; 2002US-00059911.

PR

XX (REGC ) UNIV CALIFORNIA OFFICE TECHNOLOGY.

PA

XX Fischer RL, Mizukami Y;

PI

XX WPI; 2003-787370/74.

DR

XX N-PSDB; ADG88468.

XX

XX Increasing cell proliferation in a plant by introducing into the plant an

PT expression cassette containing a plant promoter operably linked to a

PT nucleic acid encoding a modified ANT polypeptide and selecting plants

PT with increased mass.

XX

XX Claim 18; SEQ ID NO 22; Sipp; English.

PS

XX The present invention relates to plant genetic engineering. The invention

CC particularly relates to a method of increasing cell proliferation in a

CC plant. The method involves introducing into the plant an expression

CC cassette containing a plant promoter operably linked to a nucleic acid

CC encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2

CC domain and selecting plants the with increased size or mass. The

CC invention is useful to generate transgenic plants. The method is useful

CC for increasing cell proliferation in a plant for altering organ mass,

CC controlling fertility or enhancing asexual reproduction. The present

CC sequence is Arabidopsis thaliana ANT mutant protein. This sequence is

CC used in the exemplification of the invention.

XX

SQ Sequence 555 AA;

Query Match 0.8%; Score 8; DB 7; Length 555;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Search completed: August 28, 2005, 11:04:51  
Job time : 177 secs

Matches 8; :Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTL 12  
| | | | |  
Db 536 LKTVALTL 543

RESULT 100

ADG88431  
ID ADG88431 standard; protein; 555 AA.

XX AC ADG88431;

XX DT 11-MAR-2004 (first entry)

XX DE Arabidopsis thaliana aintegumenta (ANT) protein.

XX KW Plant; genetic engineering; cell proliferation; aintegumenta; ANT;  
KW transgenic; transgenic plant; organ mass alteration; fertility;  
KW asexual reproduction.

XX OS Arabidopsis thaliana.

XX FH Key Location/Qualifiers

FT Domain 281..357

FT Region /note = AP2 domain

FT FT 358..382

FT FT /note= "Linker peptide"

FT FT 383..451

FT FT /note = AP2 domain

XX PN US2003159180-A1.

XX XX 21-AUG-2003.

XX PF 28-JAN-2002; 2002US-00059911.

XX PR 28-JAN-2002; 2002US-00059911.

XX PA (REGC ) UNIV CALIFORNIA OFFICE TECHNOLOGY.

XX PI Fischer RL, Mizukami Y;

XX DR WPI; 2003-787370/74.

XX DR N-PSDB; ADG88430.

XX PT Increasing cell proliferation in a plant by introducing into the plant an  
PT expression cassette containing a plant promoter operably linked to a  
PT nucleic acid encoding a modified ANT polypeptide and selecting plants  
PT with increased mass.

XX PS Example 1; SEQ ID NO 2; 51pp; English.

XX CC The present invention relates to plant genetic engineering. The invention  
CC particularly relates to a method of increasing cell proliferation in a  
CC plant. The method involves introducing into the plant an expression  
CC cassette containing a plant promoter operably linked to a nucleic acid  
CC encoding a modified aintegumenta (ANT) polypeptide comprising an ANT AP2  
CC domain and selecting plants the with increased size or mass. The  
CC invention is useful to generate transgenic plants. The method is useful  
CC for increasing cell proliferation in a plant for altering organ mass,  
CC controlling fertility or enhancing asexual reproduction. The present  
CC sequence is Arabidopsis thaliana ANT protein.

XX SQ Sequence 555 AA;

Query Match 0.8%; Score 8; DB 7; Length 555;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTL 12  
| | | | |  
Db 536 LKTVALTL 543



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 11:08:08 ; Search time 169 Seconds

(without alignments)  
2343.607 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 1008

Sequence: 1 MKHLKTVALTITVSVVTH.....LGYTSVALLSLTAIKKKY 1008

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1767149 seqs, 392926209 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1767149

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 110 summaries

Database : Published Applications\_AA.\*

```

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1008	100.0	1008	14 US-10-078-531-2	Sequence 2, Appli
2	971	96.3	971	14 US-10-078-531-7	Sequence 7, Appli
3	907	90.0	1008	16 US-10-474-792-46	Sequence 46, Appli
4	667	66.2	969	14 US-10-078-531-8	Sequence 8, Appli
5	610	60.5	970	14 US-10-078-531-4	Sequence 4, Appli
6	540	53.6	971	14 US-10-078-531-6	Sequence 6, Appli
7	465	46.1	963	14 US-10-078-531-5	Sequence 5, Appli
8	263	26.1	951	14 US-10-078-531-3	Sequence 3, Appli
9	76	7.5	1055	14 US-10-091-007-24	Sequence 24, Appli
10	76	7.5	1055	16 US-10-476-614-2	Sequence 2, Appli
11	21	2.1	21	14 US-10-078-531-13	Sequence 13, Appli

12	9	0.9	279	16	US-10-408-765A-935	Sequence 935, App
13	9	0.9	583	15	US-10-210-281-34	Sequence 34, Appl
14	9	0.9	608	13	US-10-087-192-549	Sequence 549, App
15	9	0.9	883	15	US-10-295-027-484	Sequence 484, App
16	9	0.9	883	15	US-10-295-027-1330	Sequence 1330, Ap
17	9	0.9	883	15	US-10-173-939-28	Sequence 28, Appl
18	9	0.9	907	10	US-09-965-536A-16	Sequence 16, Appl
19	9	0.9	907	10	US-09-970-944-26	Sequence 26, Appl
20	9	0.9	907	10	US-09-970-944-26	Sequence 26, Appl
21	9	0.9	907	14	US-10-251-385-264	Sequence 264, App
22	9	0.9	907	14	US-10-251-385-278	Sequence 278, App
23	9	0.9	907	14	US-10-251-385-278	Sequence 278, App
24	9	0.9	907	15	US-10-295-027-849	Sequence 422, App
25	9	0.9	907	15	US-10-295-027-946	Sequence 849, App
26	9	0.9	907	15	US-10-295-027-1331	Sequence 946, App
27	9	0.9	907	16	US-10-751-736-84	Sequence 1331, Ap
28	9	0.9	907	17	US-10-482-029-158	Sequence 84, Appl
29	9	0.9	1145	17	US-10-505-486-103	Sequence 158, App
30	8	0.8	56	15	US-10-424-599-178222	Sequence 103, App
31	8	0.8	82	15	US-10-424-599-178222	Sequence 178222,
32	8	0.8	131	16	US-10-425-115-263516	Sequence 245882,
33	8	0.8	142	16	US-10-425-115-306321	Sequence 263516,
34	8	0.8	179	16	US-10-425-115-331445	Sequence 306321,
35	8	0.8	183	15	US-10-424-599-280877	Sequence 331445,
36	8	0.8	196	14	US-10-059-911-12	Sequence 280877,
37	8	0.8	237	17	US-10-732-923-5727	Sequence 12, Appl
38	8	0.8	239	16	US-10-425-115-262882	Sequence 5727, Ap
39	8	0.8	264	14	US-10-059-911-24	Sequence 24, Appl
40	8	0.8	275	14	US-10-059-911-25	Sequence 25, Appl
41	8	0.8	277	14	US-10-059-911-11	Sequence 11, Appl
42	8	0.8	308	14	US-10-059-911-10	Sequence 10, Appl
43	8	0.8	356	15	US-10-282-122A-45121	Sequence 45121, A
44	8	0.8	411	14	US-10-032-585-7284	Sequence 7284, Ap
45	8	0.8	458	15	US-10-369-433-17401	Sequence 17401, A
46	8	0.8	473	14	US-10-059-911-9	Sequence 9, Appli
47	8	0.8	480	18	US-10-481-032A-532	Sequence 532, App
48	8	0.8	488	15	US-10-389-566-712	Sequence 712, App
49	8	0.8	488	15	US-10-389-566-731	Sequence 731, App
50	8	0.8	488	15	US-10-389-566-1366	Sequence 1366, Ap
51	8	0.8	488	15	US-10-389-566-1367	Sequence 1367, Ap
52	8	0.8	496	18	US-10-481-032A-104	Sequence 104, App
53	8	0.8	500	15	US-10-389-566-2286	Sequence 2286, Ap
54	8	0.8	503	15	US-10-389-566-860	Sequence 860, App
55	8	0.8	503	15	US-10-389-566-1411	Sequence 1411, Ap
56	8	0.8	503	15	US-10-389-566-2188	Sequence 2188, Ap
57	8	0.8	527	15	US-10-389-566-859	Sequence 859, App
58	8	0.8	533	15	US-10-389-566-1020	Sequence 1020, Ap
59	8	0.8	533	15	US-10-389-566-1021	Sequence 1021, Ap
60	8	0.8	535	15	US-10-389-566-1041	Sequence 1041, Ap
61	8	0.8	535	15	US-10-389-566-1124	Sequence 1124, Ap
62	8	0.8	535	15	US-10-389-566-2123	Sequence 2123, Ap
63	8	0.8	535	15	US-10-389-566-2186	Sequence 2186, Ap
64	8	0.8	535	15	US-10-389-566-2397	Sequence 2397, Ap
65	8	0.8	555	14	US-10-059-911-2	Sequence 2, Appli
66	8	0.8	555	14	US-10-059-911-21	Sequence 21, Appl
67	8	0.8	555	14	US-10-059-911-22	Sequence 22, Appl
68	8	0.8	555	15	US-10-183-687-422	Sequence 422, App
69	8	0.8	555	15	US-10-183-687-424	Sequence 424, App
70	8	0.8	555	18	US-10-623-477-2	Sequence 2, Appli
71	8	0.8	642	16	US-10-739-930-6762	Sequence 6762, Ap
72	8	0.8	660	9	US-09-864-761-47959	Sequence 47959, A
73	8	0.8	686	10	US-09-981-397A-22	Sequence 22, Appl
74	8	0.8	686	18	US-10-756-149-4945	Sequence 4945, Ap
75	8	0.8	688	16	US-10-638-941-102	Sequence 102, App
76	8	0.8	715	15	US-10-425-114-54382	Sequence 54382, A
77	8	0.8	820	14	US-10-029-386-32324	Sequence 32324, A
78	8	0.8	879	17	US-10-732-923-13561	Sequence 13561, A
79	8	0.8	916	16	US-10-473-574-22	Sequence 22, Appl
80	8	0.8	931	15	US-10-170-385-39	Sequence 39, Appl
81	8	0.8	931	15	US-10-408-765A-1585	Sequence 1585, Ap
82	8	0.8	976	9	US-09-969-528-2	Sequence 2, Appli
83	8	0.8	3878	14	US-10-080-608A-11	Sequence 11, Appl
84	8	0.8	3899	14	US-10-171-311-4	Sequence 4, Appli

85 Sequence 2, Appli 8 0.8 3907 14 US-10-171-311-2  
86 Sequence 5739, Ap 8 0.8 3908 18 US-10-756-149-5739  
87 Sequence 100, App 8 0.8 3911 15 US-10-370-685-100  
88 Sequence 1839, Ap 8 0.8 3911 16 US-10-408-765A-1839  
89 Sequence 8, Appli 8 0.8 3917 14 US-10-171-311-8  
90 Sequence 6, Appli 8 0.8 3925 14 US-10-171-311-6  
91 Sequence 8, Appli 7 0.7 16 13 US-10-024-918-8  
92 Sequence 12, Appli 7 0.7 16 13 US-10-024-918-12  
93 Sequence 13, Appli 7 0.7 16 14 US-10-325-021-13  
94 Sequence 8, Appli 7 0.7 16 15 US-10-650-509-8  
95 Sequence 12, Appli 7 0.7 16 15 US-10-650-509-8  
96 Sequence 36, Appli 7 0.7 21 16 US-10-343-663A-36  
97 Sequence 45902, A 7 0.7 41 9 US-09-864-761-45902  
98 Sequence 179769, 7 0.7 46 15 US-10-424-599-179769  
99 Sequence 185551, 7 0.7 47 15 US-10-424-599-185551  
100 Sequence 225766, 7 0.7 51 15 US-10-424-599-225766  
101 Sequence 285715, 7 0.7 52 16 US-10-425-115-285715  
102 Sequence 340870, 7 0.7 52 16 US-10-425-115-340870  
103 Sequence 194696, 7 0.7 56 15 US-10-424-599-194696  
104 Sequence 247846, 7 0.7 57 15 US-10-425-115-247846  
105 Sequence 231448, 7 0.7 58 15 US-10-424-599-231448  
106 Sequence 217832, 7 0.7 63 16 US-10-425-115-217832  
107 Sequence 327358, 7 0.7 66 16 US-10-425-115-327358  
108 Sequence 364, App 7 0.7 70 9 US-09-867-550-364  
109 Sequence 249445, 7 0.7 70 16 US-10-425-115-249445  
110 Sequence 211164, 7 0.7 71 15 US-10-424-599-211164

ALIGNMENTS

RESULT 1  
US-10-078-531-2  
; Sequence 2, Application US/10078531  
; Publication No. US20030049271A1  
; GENERAL INFORMATION:  
; APPLICANT: MARTIN, DENIS  
; APPLICANT: RIOUX, STEPHANE  
; APPLICANT: BRODEUR, BERNARD R.  
; APPLICANT: HAMEL, JOSEE  
; APPLICANT: RHEAULT, PATRICK  
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING  
; TITLE OF INVENTION: DNA FRAGMENTS  
; FILE REFERENCE: PHARMA-18  
; CURRENT APPLICATION NUMBER: US/10/078,531  
; PRIOR FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: 60/269,840  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1008  
; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes  
US-10-078-531-2  
  
Query Match 100.0%; Score 1008; DB 14; Length 1008;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MKKHLKVALTITVSVTHNQEVSLVKEPIKQTOASSISGADYAESGSKLKINE 60  
DB 1 MKKHLKVALTITVSVTHNQEVSLVKEPIKQTOASSISGADYAESGSKLKINE 60  
  
QY 61 TSGPVDVDTTDLFSDKRTTPEKIKNLAKGPREQELKAVTENSEKQITSGSQLEQSK 120  
DB 61 TSGPVDVDTTDLFSDKRTTPEKIKNLAKGPREQELKAVTENSEKQITSGSQLEQSK 120  
  
QY 121 SLSLNKTVPSTNWEICDFTKGTNTLVGLSKGVEKLSQTDHLVLPQAADGTQLIQVAS 180  
DB 121 SLSLNKTVPSTNWEICDFTKGTNTLVGLSKGVEKLSQTDHLVLPQAADGTQLIQVAS 180  
  
QY 181 FAFTPDKKTAAEYTSRAGENGESQLDVGKEIINEGEVFNLSLLKVTIPTGYKHIGQ 240

181 FAFTPDKKTAAEYTSRAGENGESQLDVGKEIINEGEVFNLSLLKVTIPTGYKHIGQ 240  
QY 241 DAFVNDKNIAEVLNIPESLETISDYAFALHALKQDLDPNLKAIGELAFDFOIITGKLSLP 300  
DB 241 DAFVNDKNIAEVLNIPESLETISDYAFALHALKQDLDPNLKAIGELAFDFOIITGKLSLP 300  
QY 301 ROLARLAEAPKSNHKTIEFRGNSLVKIGRASQDNDLSQMLPDGKLEKTESAFTGNP 360  
DB 301 ROLARLAEAPKSNHKTIEFRGNSLVKIGRASQDNDLSQMLPDGKLEKTESAFTGNP 360  
QY 361 GDDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLQWESPEIDYTKWLEEDFTYQKNSVT 420  
DB 361 GDDHNNRVVLTWTKSGKNPSGLATENTYVNPDKSLQWESPEIDYTKWLEEDFTYQKNSVT 420  
QY 421 GFSNKGLOKVRKNLNIPIKOHNGVTITEIGDGNFRVNDPONTLRKYDLEBEVKLPSTIR 480  
DB 421 GFSNKGLOKVRKNLNIPIKOHNGVTITEIGDGNFRVNDPONTLRKYDLEBEVKLPSTIR 480  
QY 481 KIGAPAFQSNLKSFEASDDLEEKEGAFMNNRLETLELKDQVITIGDAAPHINHIYAIV 540  
DB 481 KIGAPAFQSNLKSFEASDDLEEKEGAFMNNRLETLELKDQVITIGDAAPHINHIYAIV 540  
QY 541 LPESVQIEGRSAFRQNGANNLIFMGSKVKTLGEMAFNLSNRLEHLDLSEKQLTPIVQAF 600  
DB 541 LPESVQIEGRSAFRQNGANNLIFMGSKVKTLGEMAFNLSNRLEHLDLSEKQLTPIVQAF 600  
QY 601 SDNALKEVLLPASLKTITREEAFKKNHLKQLEVASALSIAFNALDDNDGDSQFQNKVVVK 660  
DB 601 SDNALKEVLLPASLKTITREEAFKKNHLKQLEVASALSIAFNALDDNDGDSQFQNKVVVK 660  
QY 661 THNSYALADGEHFIIVDPDKLSSTIVDLKTLKLEGLDYSTLRQTQTQFRDMTTAGKA 720  
DB 661 THNSYALADGEHFIIVDPDKLSSTIVDLKTLKLEGLDYSTLRQTQTQFRDMTTAGKA 720  
QY 721 LLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKATKNGOLLERSINKAVL 780  
DB 721 LLSKSNLRQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKATKNGOLLERSINKAVL 780  
QY 781 AYNSAIKANVKKRLEKELDLTLGLVEGKGLAQATWVGYYLLKTPPLPEYIYGLNVY 840  
DB 781 AYNSAIKANVKKRLEKELDLTLGLVEGKGLAQATWVGYYLLKTPPLPEYIYGLNVY 840  
QY 841 FDKSGKLIYALDMSDITGEGQKDAYGNPILNVDEGEGYHALAVATLADYEGLDIKTILN 900  
DB 841 FDKSGKLIYALDMSDITGEGQKDAYGNPILNVDEGEGYHALAVATLADYEGLDIKTILN 900  
QY 901 SKLSQLTSIROVPTAAVHRAGIFQAIQNAAEASQLLPKQTHSEKSSSSSSANSKDRGL 960  
DB 901 SKLSQLTSIROVPTAAVHRAGIFQAIQNAAEASQLLPKQTHSEKSSSSSSANSKDRGL 960  
QY 961 QSNPKTNRGRHSAILPRTGSKGSFVYIGLYTSVALLSLITAIKKKKY 1008  
DB 961 QSNPKTNRGRHSAILPRTGSKGSFVYIGLYTSVALLSLITAIKKKKY 1008

RESULT 2  
US-10-078-531-7  
; Sequence 7, Application US/10078531  
; Publication No. US20030049271A1  
; GENERAL INFORMATION:  
; APPLICANT: MARTIN, DENIS  
; APPLICANT: RIOUX, STEPHANE  
; APPLICANT: BRODEUR, BERNARD R.  
; APPLICANT: HAMEL, JOSEE  
; APPLICANT: RHEAULT, PATRICK  
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING  
; TITLE OF INVENTION: DNA FRAGMENTS  
; FILE REFERENCE: PHARMA-18  
; CURRENT APPLICATION NUMBER: US/10/078,531  
; PRIOR FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: 60/269,840  
; PRIOR FILING DATE: 2001-02-21

; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 971  
; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes  
US-10-078-531-7

```
Query Match      96.3%; Score 971; DB 14; Length 971;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 LVKEPILKQTOASSISGADYAESGSKLKINETSGPVDVDTDLFSDKRTTPEKIKON 86
DB 1 LVKEPILKQTOASSISGADYAESGSKLKINETSGPVDVDTDLFSDKRTTPEKIKON 60
QY 87 LAKGPREQELKAVTENTSEKQITSGSQLEQESLSLNKTVSTNWEICDPITKGNL 146
DB 61 LAKGPREQELKAVTENTSEKQITSGSQLEQESLSLNKTVSTNWEICDPITKGNL 120
QY 147 VGLSGSVEKLSQTDHLVLPQAAADGTQLIQVAFAPFTDPDKTAIAEYTSRAGENGESIQ 206
DB 121 VGLSGSVEKLSQTDHLVLPQAAADGTQLIQVAFAPFTDPDKTAIAEYTSRAGENGESIQ 180
QY 207 LDVDGKEIINEGVFNFSYLLKKVTIPTGYKHIGQDAFVDMKNIAEVLNLPESLETISDYAF 266
DB 181 LDVDGKEIINEGVFNFSYLLKKVTIPTGYKHIGQDAFVDMKNIAEVLNLPESLETISDYAF 240
QY 267 AHLALAKQIDLPDLNKAIGELAFPDNQITGKLSLPRQLMRLAERAFKSNHKTIEFRGNSL 326
DB 241 AHLALAKQIDLPDLNKAIGELAFPDNQITGKLSLPRQLMRLAERAFKSNHKTIEFRGNSL 300
QY 327 KVTGEASFQDNDLSQMLPDGLSEKIESEAFNGPDHNNRVVLTWTKSGKNPSGLATEN 386
DB 301 KVTGEASFQDNDLSQMLPDGLSEKIESEAFNGPDHNNRVVLTWTKSGKNPSGLATEN 360
QY 387 TYNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPKQNGVT 446
DB 361 TYNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKVRKNKLEIPKQNGVT 420
QY 447 ITEIGNAPRVDFQNTLRYDLBEVKLPSTTRKIGAFAPQSNLKSFEASDDLEIYE 506
DB 421 ITEIGNAPRVDFQNTLRYDLBEVKLPSTTRKIGAFAPQSNLKSFEASDDLEIYE 480
QY 507 GAPMNNRIETLEKDKLVITGDAAFINHIYATVLPESVQEIERSAFRONGANNLIFMGS 566
DB 481 GAPMNNRIETLEKDKLVITGDAAFINHIYATVLPESVQEIERSAFRONGANNLIFMGS 540
QY 567 KVKTGEMAFSLNRLHDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREAFKKNH 626
DB 541 KVKTGEMAFSLNRLHDLSEKQLTEIPVQAFSDNALKEVLLPASLKTIREAFKKNH 600
QY 627 LKQLEVASALSHITAFNALDNDGDEQFNDKVVVTHNSYALADGSHFIVDPDKLSSTIV 686
DB 601 LKQLEVASALSHITAFNALDNDGDEQFNDKVVVTHNSYALADGSHFIVDPDKLSSTIV 660
QY 687 DLEKILKLEGLDYSTLROTTQFQDMTTFAGKALLSKNLRQGEKQKFOEAQPELGRV 746
DB 661 DLEKILKLEGLDYSTLROTTQFQDMTTFAGKALLSKNLRQGEKQKFOEAQPELGRV 720
QY 747 DLDKATAKAEKALVTKATQNGOLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLV 806
DB 721 DLDKATAKAEKALVTKATQNGOLLERSINKAVLAYNNSAIKKANVKRLEKELDLTLGLV 780
QY 807 EGKGPLAQATWQGVYLLKTPPLPEYIYGLNVPYFDKSGKLIYALDMSDTIGSGKDAYG 866
DB 781 EGKGPLAQATWQGVYLLKTPPLPEYIYGLNVPYFDKSGKLIYALDMSDTIGSGKDAYG 840
QY 867 NPILANVEDNEGHALAVATLADYEGLDIKTILNSKLSQTSIRQVPTAAYHRAFIQAI 926
DB 841 NPILANVEDNEGHALAVATLADYEGLDIKTILNSKLSQTSIRQVPTAAYHRAFIQAI 900
QY 927 QNAAAEAEQLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKGSFVY 986
```

```
DB 901 QNAAAEAEQLPKPGTHSEKSSSESANSKDRGLQSNPKTNRGRHSAILPRTGSKGSFVY 960
QY 987 GILGTVSVAL 997
DB 961 GILGTVSVAL 971

RESULT 3
US-10-474-792-46
; Sequence 46, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Winter, Laurie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474,792
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 46
; LENGTH: 1008
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-474-792-46
```

Query Match 90.0%; Score 907; DB 16; Length 1008;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1007; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 MKKHLKTVALTITTVSVVTHNQEVSLVKEPILKQTOASSISGADYAESGSKLKINE 60
DB 1 MKKHLKTVALTITTVSVVTHNQEVSLVKEPILKQTOASSISGADYAESGSKLKINE 60
QY 61 TSGPVDVDTVDLFSKRTTPEKIKONLAGPREQELKAVTENTSEKQITSGSQLEQSK 120
DB 61 TSGPVDVDTVDLFSKRTTPEKIKONLAGPREQELKAVTENTSEKQITSGSQLEQSK 120
QY 121 SLSLNTKVPSTNWEICDPITKGNLTVGLSGSVEKLSQTDHLVLPQAAADGTQLIQVAS 180
DB 121 SLSLNTKVPSTNWEICDPITKGNLTVGLSGSVEKLSQTDHLVLPQAAADGTQLIQVAS 180
QY 181 FAPTPDKKTAIAEYTSRAGENGESIQLDVDGKEIINEGVFNFSYLLKKVTIPTGYKHIG 240
DB 181 FAPTPDKKTAIAEYTSRAGENGESIQLDVDGKEIINEGVFNFSYLLKKVTIPTGYKHIG 240
QY 241 DAFVDMKNIAEVLNLPESLETISDYAFAPHLAKQIDLPDLNKAIGELAPFDNQITGKLSLP 300
DB 241 DAFVDMKNIAEVLNLPESLETISDYAFAPHLAKQIDLPDLNKAIGELAPFDNQITGKLSLP 300
QY 301 RQMLRAERAFKSNHKTIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEAFNGP 360
DB 301 RQMLRAERAFKSNHKTIEFRGNSLKVIGEASFQDNDLSQMLPDGLEKIESEAFNGP 360
QY 361 GDHYNRRVLTWTKSGKNPSGLATENTYVNDPKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
DB 361 GDHYNRRVLTWTKSGKNPSGLATENTYVNDPKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
QY 421 GFSNKGLOKVRKNKLEIPKQNGVTITTEIGNAPRVDFQNTLRYDLBEVKLPSTTR 480
DB 421 GFSNKGLOKVRKNKLEIPKQNGVTITTEIGNAPRVDFQNTLRYDLBEVKLPSTTR 480
QY 481 KIGAFAPQSNLKSFEASDDLEEKEGAFMNNRIETLEKDKLVITGDAAFINHIYATV 540
DB 481 KIGAFAPQSNLKSFEASDDLEEKEGAFMNNRIETLEKDKLVITGDAAFINHIYATV 540
QY 541 LPESVQEIERSAFRONGANNLIFMGSVKVTKLGEMAFSLNRLHDLSEKQLTEIPVQAF 600
DB 541 LPESVQEIERSAFRONGANNLIFMGSVKVTKLGEMAFSLNRLHDLSEKQLTEIPVQAF 600
```

QY 601 SDNALKEVLLPASLTIREAPKKNHLKQLEVASALSHTAFNALDDNDGDEQFNKVYK 660  
DB 601 SDNALKEVLLPASLTIREAPKKNHLKQLEVASALSHTAFNALDDNDGDEQFNKVYK 660  
QY 661 THNSYALADGEHFVDPDKLSTTIVLEKILKLEGLDYSLRQTQTQFRDMTTAGKA 720  
DB 661 THNSYALADGEHFVDPDKLSTTIVLEKILKLEGLDYSLRQTQTQFRDMTTAGKA 720  
QY 721 LLSKNLROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKATKNGOLLERSINKAVL 780  
DB 721 LLSKNLROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKATKNGOLLERSINKAVL 780  
QY 781 AYNSAIKANKYKLEKELDLTLGLVEGKPLAQATMVGVLKATPLPEYIIGLVNY 840  
DB 781 AYNSAIKANKYKLEKELDLTLGLVEGKPLAQATMVGVLKATPLPEYIIGLVNY 840  
QY 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDEGYPHALAVATLADYEGLDIKTILN 900  
DB 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDEGYPHALAVATLADYEGLDIKTILN 900  
QY 901 SKLSQTSIRQVPTAAYHRAGIFQAIQNAAAAEQOLLPKPGTHSEKSSSESANSKDRGL 960  
DB 901 SKLSQTSIRQVPTAAYHRAGIFQAIQNAAAAEQOLLPKPGTHSEKSSSESANSKDRGL 960  
QY 961 QSNPKTNRGRHSAILPRTGSKGSFVYGLGYTSVALLSLITAIKKKKY 1008  
DB 961 QSNPKTNRGRHSAILPRTGSKGSFVYGLGYTSVALLSLITAIKKKKY 1008

## RESULT 4

US-10-078-531-8  
; Sequence 8, Application US/10078531  
; Publication No. US20030049271A1  
; GENERAL INFORMATION:  
; APPLICANT: MARTIN, DENIS  
; APPLICANT: RIOUX, STEPHANE  
; APPLICANT: BRODEUR, BERNARD R.  
; APPLICANT: HAMEL, JOSEE  
; APPLICANT: RHEAULT, PATRICK  
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING  
; FILE REFERENCE: PHARMA-18  
; CURRENT APPLICATION NUMBER: US/10/078,531  
; PRIOR FILING DATE: 2002-02-21  
; PRIOR FILING DATE: 2001-02-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 969  
; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes  
US-10-078-531-8

Query Match 66.2%; Score 667; DB 14; Length 969;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 867; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 128 VPSTSNWEICDITTKGNTLVGLSGVKEKLSQTDHLVLPQAAQDTQLIQVASFAPTPDK 187  
DB 101 VPSTSNWEICDITTKGNTLVGLSGVKEKLSQTDHLVLPQAAQDTQLIQVASFAPTPDK 160  
QY 188 KTAIAEYTSRAGENGESOLDVDGKEIINEGVFNSYLLKKVTIPTGYKHIGQDAFVNK 247  
DB 161 KTAIAEYTSRAGENGESOLDVDGKEIINEGVFNSYLLKKVTIPTGYKHIGQDAFVNK 220  
QY 248 NTAENVLPESLTIIDYAPAHALAKQIDLDPNLKKAIGELAFPDNQITGKLSLPROLMELA 307  
DB 221 NTAENVLPESLTIIDYAPAHALAKQIDLDPNLKKAIGELAFPDNQITGKLSLPROLMELA 280  
QY 308 ERAFKSNHIKTIEFRGNSLKVTIGEASFQDNDLSQLMLPDGLEKIESEFTGNPGDDHNN 367  
DB 281 ERAFKSNHIKTIEFRGNSLKVTIGEASFQDNDLSQLMLPDGLEKIESEFTGNPGDDHNN 340

QY 368 RVVLWTGSKGNPSGLATENTYVNPDKSLWQSSPEIDYTKMLBEDFTYQKNSVTGFSNKL 427  
DB 341 RVVLWTGSKGNPSGLATENTYVNPDKSLWQSSPEIDYTKMLBEDFTYQKNSVTGFSNKL 400  
QY 428 QKVRKNKLEIPKOHNGVTITEIGDAPRNVDFONKTLRKYDLEEVKLPSITIRKIGAFAP 487  
DB 401 QKVRKNKLEIPKOHNGVTITEIGDAPRNVDFONKTLRKYDLEEVKLPSITIRKIGAFAP 460  
QY 488 QSNLKSFEASDDLEEKEGAPMNNRIETLEKDKLVITIGDAAPHINHIYAIVLPEVQE 547  
DB 461 QSNLKSFEASDDLEEKEGAPMNNRIETLEKDKLVITIGDAAPHINHIYAIVLPEVQE 520  
QY 548 IGRSAFRONGANNLIFMGSKVKTIGEMAFISNRLEHLDLSEKQKLTETIPVQAFSDNALKE 607  
DB 521 IGRSAFRONGANNLIFMGSKVKTIGEMAFISNRLEHLDLSEKQKLTETIPVQAFSDNALKE 580  
QY 608 VLLPASLKTIREEAPKKNHLKQLEVASALSHTAFNALDDNDGDEQFNKVYKTHNSYA 667  
DB 591 VLLPASLKTIREEAPKKNHLKQLEVASALSHTAFNALDDNDGDEQFNKVYKTHNSYA 640  
QY 668 LADGEHFIVDPDKLSTTIVLEKILKLEGLDYSLRQTQTQFRDMTTAGKALLSKSNL 727  
DB 641 LADGEHFIVDPDKLSTTIVLEKILKLEGLDYSLRQTQTQFRDMTTAGKALLSKSNL 700  
QY 728 RQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKATKNGOLLERSINKAVLAYNNSAI 787  
DB 701 RQGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKKATKNGOLLERSINKAVLAYNNSAI 760  
QY 788 KKANVKRLEKELDLTLGLVEGKPLAQATMVGVLKATPLPEYIIGLVNYFDKSGKL 847  
DB 761 KKANVKRLEKELDLTLGLVEGKPLAQATMVGVLKATPLPEYIIGLVNYFDKSGKL 820  
QY 848 IYALDMSDTIGEGQKDAYGNPILNVDEDEGYPHALAVATLADYEGLDIKTILSKLSQLT 907  
DB 821 IYALDMSDTIGEGQKDAYGNPILNVDEDEGYPHALAVATLADYEGLDIKTILSKLSQLT 880  
QY 908 SIRQVPTAAYHRAGIFQAIQNAAAAEQOLLPKPGTHSEKSSSESANSKDRGLQSNPKTN 967  
DB 881 SIRQVPTAAYHRAGIFQAIQNAAAAEQOLLPKPGTHSEKSSSESANSKDRGLQSNPKTN 940  
QY 968 RGRHSAILPRTGSKGSFVYGLGYTSVAL 996  
DB 941 RGRHSAILPRTGSKGSFVYGLGYTSVAL 969

## RESULT 5

US-10-078-531-4  
; Sequence 4, Application US/10078531  
; Publication No. US20030049271A1  
; GENERAL INFORMATION:  
; APPLICANT: MARTIN, DENIS  
; APPLICANT: RIOUX, STEPHANE  
; APPLICANT: BRODEUR, BERNARD R.  
; APPLICANT: HAMEL, JOSEE  
; APPLICANT: RHEAULT, PATRICK  
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING  
; FILE REFERENCE: PHARMA-18  
; CURRENT APPLICATION NUMBER: US/10/078,531  
; PRIOR FILING DATE: 2002-02-21  
; PRIOR FILING DATE: 2001-02-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 970  
; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes  
US-10-078-531-4

Query Match 60.5%; Score 610; DB 14; Length 970;  
Best Local Similarity 99.8%; Pred. No. 0;

Matches	810;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
Qy	128	VPSTSNWEICDPIITKGNLTVLGSKSGVEKLSQTDHVLPLPSQAADGTQLIQVASFAFTPDK	187						
Db	102	VPSTSNWEICDPIITKGNLTVLGSKSGVEKLSQTDHVLPLPSQAADGTQLIQVASFAFTPDK	161						
Qy	188	KTATAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTTPTGYKHIGQDAFVDNK	247						
Db	162	KTATAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTTPTGYKHIGQDAFVDNK	221						
Qy	248	NIAEVNLPESLETTSDYAFAHALAKQIDLPNLKAIAGELAFDFDNOITGKLSLPQLMRLA	307						
Db	222	NIAEVNLPESLETTSDYAFAHALAKQIDLPNLKAIAGELAFDFDNOITGKLSLPQLMRLA	281						
Qy	308	ERAFKSNHIIKTIIEPRGNSLKVIGBASFOONDLSQMLPDLGLEKIESEAFTPNGDDHYNH	367						
Db	282	ERAFKSNHIIKTIIEPRGNSLKVIGBASFOONDLSQMLPDLGLEKIESEAFTPNGDDHYNH	341						
Qy	368	RVVLWTSGKNPGLATENTVVPDKSLUQWESPEIDYTKMLDEEFTYQKNSVTGSPNKL	427						
Db	342	RVVLWTSGKNPGLATENTVVPDKSLUQWESPEIDYTKMLDEEFTYQKNSVTGSPNKL	401						
Qy	428	QKVGRNKNLEIPKQHGNGVTITEIGDNAFRNVDFOFKTLRKDYDLSEVKLPSTIRKIGAFAP	487						
Db	402	QKVGRNKNLEIPKQHGNGVTITEIGDNAFRNVDFOFKTLRKDYDLSEVKLPSTIRKIGAFAP	461						
Qy	488	QSNNLKSFASDDLEEIKEGAFMNRITETLELKKLVITIGDAAFHINHHIYAIVLPESVOE	547						
Db	462	QSNNLKSFASDDLEEIKEGAFMNRITETLELKKLVITIGDAAFHINHHIYAIVLPESVOE	521						
Qy	548	IGRSAPFRQNGANNLIFMGSKVKTIGEMAFNSNRLEHLDLSQKOLTEIPVOAFSDNALKE	607						
Db	522	IGRSAPFRQNGANNLIFMGSKVKTIGEMAFNSNRLEHLDLSQKOLTEIPVOAFSDNALKE	581						
Qy	608	VLLPASLKTIREEAPKQNHKLQOLEVASALSHIAFNALDDNDGDSEQFONKVVVVKTHNSYA	667						
Db	582	VLLPASLKTIREEAPKQNHKLQOLEVASALSHIAFNALDDNDGDSEQFONKVVVVKTHNSYA	641						
Qy	668	LADGEHFTVPDKLSSTIVDLEKILKULEGHLDYSTLRQTTOTQPRDMTTAGKALLSKNL	727						
Db	642	LADGEHFTVPDKLSSTIVDLEKILKULEGHLDYSTLRQTTOTQPRDMTTAGKALLSKNL	701						
Qy	728	RQGEKQFLQEAQOFLGRVDLDKAIAKAEKALVTKKATKNGQLLERSINKAVLAYNNSAI	787						
Db	702	RQGEKQFLQEAQOFLGRVDLDKAIAKAEKALVTKKATKNGQLLERSINKAVLAYNNSAI	761						
Qy	788	KKANVKLEKELDLTLTGLVEKGPLAQATWQGVYLLKTPPLPBPYYIGLNVYFPKSGKL	847						
Db	762	KKANVKLEKELDLTLTGLVEKGPLAQATWQGVYLLKTPPLPBPYYIGLNVYFPKSGKL	821						
Qy	848	IYALDMSDTTIGEGOKDAYGNPILNVDEDEGPHALAVATTLADYEGLDIKTILNSKLSOLT	907						
Db	822	IYALDMSDTTIGEGOKDAYGNPILNVDEDEGPHALAVATTLADYEGLDIKTILNSKLSOLT	881						
Qy	908	SIROVPTAAHYHRAGIFQAIQNAABAEQLLPK	939						
Db	882	SIROVPTAAHYHRAGIFQAIQNAABAEQLLPK	913						

```

RESULT 6
US-10-078-531-6
; Sequence 6, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEF
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078.531

```

;	CURRENT FILING DATE: 2002-02-21	
;	PRIOR APPLICATION NUMBER: 60/269,840	
;	PRIOR FILING DATE: 2001-02-21	
;	NUMBER OF SEQ ID NOS: 14	
;	SOFTWARE: PatentIn Ver. 2.1	
;	SEQ ID NO 6	
;	LENGTH: 971	
;	TYPE: PRT	
;	ORGANISM: Streptococcus pyogenes	
US-10-078-531-6		

  

Query Match	53.6%;	Score 540;	DB 14;	Length 971;
Best Local Similarity	99.6%;	Pred. No. 0;		
Matches 840;	Conservative 0;	Mismatches 3;	Indels	

  

Qy	99	VTENTESEKQITSGSQLEQSEKESLSLNKNTVPSTSNWEICDPITKGNLTAVGL		
Db	73	VTENTESEKQITSGSQLEQSEKESLSLNKNTVPSTSNWEICDPITKGNLTAVGL		
Qy	159	QTDHLVLPQAADGTQLIQVASFAFTPDKKTAIAEYTSRAGENGESIQOLDV		
Db	133	QTDHLVLPQAADGTQLIQVASFAFTPDKKTAIAEYTSRAGENGESIQOLDV		
Qy	219	EVFNSYLLKKVTPTYGKYKHIGQDAFVGNKNIAEVLNPESLETISDYAFAH		
Db	193	EVFNSYLLKKVTPTYGKYKHIGQDAFVGNKNIAEVLNPESLETISDYAFAH		
Qy	279	NLKAIGELAFPDNQITCKLSIPQLMELAEAPFKSNHIIKTIETPRGNSLKV		
Db	253	NLKAIGELAFPDNQITCKLSIPQLMELAEAPFKSNHIIKTIETPRGNSLKV		
Qy	339	LSQLMLPDGLEKIESEAFNGPGDDHYNRRVVLMTKSGKNPSGLATENTVY		
Db	313	LSQLMLPDGLEKIESEAFNGPGDDHYNRRVVLMTKSGKNPYGLATENTVY		
Qy	399	SPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKQKRNKNLLEIPKOHNGVTIT		
Db	373	SPEIDYTKWLEEDFTYQKNSVTGFSNKGLOKQKRNKNLLEIPKOHNGVTIT		
Qy	459	DFQNKTLRKDYDLBEVKLPSTIRKIGAFAPFOSNNLKSPASDDLEEKEGAR		
Db	433	DFQNKTLRKDYDLBEVKLPSTIRKIGAFAPFOSNNLKSPASDDLEEKEGAR		
Qy	519	LKDKLVTIGDAAPHINHIYAILVPESVQETGRSAFRONGANNILFMGSKVKY		
Db	493	LKDKLVTIGDAAPHINHIYAILVPESVQETGRSAFRONGANNILFMGSKVKY		
Qy	579	NRLEHLDLSEQQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKGNHLKQ		
Db	553	NRLEHLDLSEQQLTEIPVQAFSDNALKEVLLPASLKTIREEAFKGNHLKQ		
Qy	639	IAFNALDDNDGDDQFQFNKVVVKTTHNSYALADGEHFTVDPDKLSSTIVDLE		
Db	613	IAFNALDDNDGDDQFQFNKVVVKTTHNSYALADGEHFTVDPDKLSSTIVDLE		
Qy	699	DYSTLRQTTQTQFRDMTTAGKALLSKNSLRQSEKQKFLQEAQFPLGRVDLDD		
Db	673	DYSTLRQTTQTQFRDMTTAGKALLSKNSLRQSEKQKFLQEAQFPLGRVDLDD		
Qy	759	LVTKKATKNGQLLERSINKAVLAYNNSAIKKANVYKRLKEKELDLLTGLVGE		
Db	733	LVTKKATKNGQLLERSINKAVLAYNNSAIKKANVYKRLKEKELDLLTGLVGE		
Qy	819	QGVYLLKTPLPPEYYIIGLVNVPFDKSKGLIYALDMSDTIGEGQKDAYGNPI		
Db	793	QGVYLLKTPLPPEYYIIGLVNVPFDKSKGLIYALDMSDTIGEGQKDAYGNPI		
Qy	879	YHALAVATTLADYEGLDIKTILNSKLSQLTISIROPVTAAYHRAGIFQAOIQA		
Db	853	YHALAVATTLADYEGLDIKTILNSKLSQLTISIROPVTAAYHRAGIFQAOIQA		
Qy	939	KPG 941		

```
Db      913 KPG 915

RESULT 7
US-10-078-531-5
; Sequence 5, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; LENGTH: 963
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-5

Query Match      46.1%; Score 465; DB 14; Length 963;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 865; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 128 VPSTSNWEICDPTITKGNLTVLGSLKSGVEKLSQTDHLVLPQQAADGTQLIQVAFPTDPK 187
Db      95 VPSTSNWEICDPTITKGNLTVLGSLKSGVEKLSQTDHLVLPQQAADGTQLIQVAFPTDPK 154
QY 188 KTAIAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNK 247
Db      155 KTAIAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNK 214
QY 248 NIAEVNLPESLETISDYAFALHALKQIDLDPNLKAIAGELAFPDNQITGKLSLPRQLMRLA 307
Db      215 NIAEVNLPESLETISDYAFALHALKQIDLDPNLKAIAGELAFPDNQITGKLSLPRQLMRLA 274
QY 308 ERAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQLMPLDGLKIESEAFPTGNPGDDHNN 367
Db      275 ERAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQLMPLDGLKIESEAFPTGNPGDDHNN 334
QY 368 RVVLTWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKG 427
Db      335 RVVLTWTKSGKNPYGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKG 394
QY 428 QKVAKNKLKLEIPKQHGVTITEIGDNAPRNVDVFNKTLRKYLDEEVKLPSTIRKIGAFAP 487
Db      395 QKVAKNKLKLEIPKQHGVTITEIGDNAPRNVDVFNKTLRKYLDEEVKLPSTIRKIGAFAP 454
QY 488 QSNNLKSFASDDLEIEKEGAFMNNRIETLEKDKLVITIGDAAFPHNHIYAIVLPSVQOE 547
Db      455 QSNNLKSFASDDLEIEKEGAFMNNRIETLEKDKLVITIGDAAFPHNHIYAIVLPSVQOE 514
QY 548 IGRSAPFRONGANNLIIFMGSKVTLGEMAFSLNRLHLDLSEOKLTETIPVQAFSNALKE 607
Db      515 IGRSAPFRONGANNLIIFMGSKVTLGEMAFSLNRLHLDLSEOKLTETIPVQAFSNALKE 574
QY 608 VLLPASLKTIREAFKONHLKQLEVASALSHAFNALDNDGDEQDNKVVVKTTHNSYA 667
Db      575 VLLPASLKTIREAFKONHLKQLEVASALSHAFNALDNDGDEQDNKVVVKTTHNSYA 634
QY 668 LADGSHFIVDPDKLSSTIVDLKILKLIBGLDYSLRQTTQTPQFRDWTAGKALLSKSNL 727
Db      635 LADGSHFIVDPDKLSSTIVDLKILKLIBGLDYSLRQTTQTPQFRDWTAGKALLSKSNL 694
QY 728 ROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKATNGQLLERSINKAVIAYNNSAI 787

Db      695 ROGEKQKFLQEAQFFLGRVLDLKAIAKAEKALVTKATNGQLLERSINKAVIAYNNSAI 754
QY 788 KKANVKRLKELDLTLGLVEGKGPLAQATWVQGVYLLKTLPLPEYIYIGLVNVPDKSGKL 847
Db      755 KKANVKRLKELDLTLGLVEGKGPLAQATWVQGVYLLKTLPLPEYIYIGLVNVPDKSGKL 814
QY 848 IYALDMSDTIGEGQKDAYGNPILNVDSBNBYHALAVATLADYEGLODKITLNSKLSQUT 907
Db      815 IYALDMSDTIGEGQKDAYGNPILNVDSBNBYHALAVATLADYEGLODKITLNSKLSQUT 874
QY 908 SIROVPTAAVHRAGIFQAIQNAABAEQQLLPKPGTHSEKSSSSSANSKORGLQSNPKTN 967
Db      875 SIROVPTAAVHRAGIFQAIQNAABAEQQLLPKPGTHSEKSSSSSANSKORGLQSNPKTN 934
QY 968 RGRHSAILPRTGSKGSFVYIGLYTSVAL 996
Db      935 RGRHSAILPRTGSKGSFVYIGLYTSVAL 963

RESULT 8
US-10-078-531-3
; Sequence 3, Application US/10078531
; Publication No. US20030049271A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, DENIS
; APPLICANT: RIOUX, STEPHANE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: HAMEL, JOSEE
; APPLICANT: RHEAULT, PATRICK
; TITLE OF INVENTION: STREPTOCOCCUS PYOGENES POLYPEPTIDES AND CORRESPONDING
; FILE REFERENCE: PHARMA-18
; CURRENT APPLICATION NUMBER: US/10/078,531
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/269,840
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 951
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-078-531-3

Query Match      26.1%; Score 263; DB 14; Length 951;
Best Local Similarity 99.3%; Pred. No. 1.8e-243;
Matches 863; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 128 VPSTSNWEICDPTITKGNLTVLGSLKSGVEKLSQTDHLVLPQQAADGTQLIQVAFPTDPK 187
Db      83 VPSTSNWEICDPTITKGNLTVLGSLKSGVEKLSQTDHLVLPQQAADGTQLIQVAFPTDPK 142
QY 188 KTAIAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNK 247
Db      143 KTAIAEYTSRAGENGESOLDVDGKEIINEGEVFNYSLLKKVTIPTGYKHIGQDAFVDNK 202
QY 248 NIAEVNLPESLETISDYAFALHALKQIDLDPNLKAIAGELAFPDNQITGKLSLPRQLMRLA 307
Db      203 NIAEVNLPESLETISDYAFALHALKQIDLDPNLKAIAGELAFPDNQITGKLSLPRQLMRLA 262
QY 308 ERAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQLMPLDGLKIESEAFPTGNPGDDHNN 367
Db      263 ERAFKSNHIKTIIEFRGNLSKVIGEASFQDNDLSQLMPLDGLKIESEAFPTGNPGDDHNN 322
QY 368 RVVLTWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKG 427
Db      323 RVVLTWTKSGKNPYGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVTGFSNKG 382
QY 428 QKVAKNKLKLEIPKQHGVTITEIGDNAPRNVDVFNKTLRKYLDEEVKLPSTIRKIGAFAP 487
Db      383 QKVAKNKLKLEIPKQHGVTITEIGDNAPRNVDVFNKTLRKYLDEEVKLPSTIRKIGAFAP 442
```





; APPLICANT: Ghosh, Soumitra S.  
 ; APPLICANT: Faby, Eoin D.  
 ; APPLICANT: Zhang, Bing  
 ; APPLICANT: Gibson, Bradford W.  
 ; APPLICANT: Taylor, Steven W.  
 ; APPLICANT: Glenn, Gary M.  
 ; APPLICANT: Warnock, Dale E.  
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
 ; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
 ; FILE REFERENCE: 660088.465  
 ; CURRENT APPLICATION NUMBER: US/10/408,765A  
 ; CURRENT FILING DATE: 2003-04-04  
 ; NUMBER OF SEQ ID NOS: 3077  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 935  
 ; LENGTH: 279  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-408-765A-935

Query Match 0.9%; Score 9; DB 16; Length 279;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSES 952  
 Db 18 SEKSSSES 26

RESULT 13  
 US-10-210-281-34  
 ; Sequence 34, Application US/10210281  
 ; Publication No. US20040030096A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gorman, Linda  
 ; APPLICANT: Zethusen, Bryan D.  
 ; APPLICANT: Edinger, Shlomit R.  
 ; APPLICANT: Padigaru, Muralidhara  
 ; APPLICANT: Guo, Xiaojia  
 ; APPLICANT: Kekuda, Ramesh  
 ; APPLICANT: Zhong, Mei  
 ; APPLICANT: Patturajan, Meera  
 ; APPLICANT: Miller, Charles E.  
 ; APPLICANT: Jij, Weizhen  
 ; APPLICANT: Pena, Carol E.A.  
 ; APPLICANT: Bugesse, Catherine E.  
 ; APPLICANT: Sciore, Paul  
 ; APPLICANT: Stone, David J.  
 ; APPLICANT: Taupier, Raymond J., Jr.  
 ; APPLICANT: Cahan, Stacie  
 ; APPLICANT: Rothenberg, Mark E.  
 ; APPLICANT: Malyankar, Uriel M.  
 ; APPLICANT: Boldog, Ferenc L.  
 ; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS  
 ; FILE REFERENCE: 21402-416D  
 ; CURRENT APPLICATION NUMBER: US/10/210,281  
 ; CURRENT FILING DATE: 2003-02-05  
 ; PRIOR APPLICATION NUMBER: 60/309,501  
 ; PRIOR FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 60/310,291  
 ; PRIOR FILING DATE: 2001-08-03  
 ; PRIOR APPLICATION NUMBER: 60/361,775  
 ; PRIOR FILING DATE: 2002-03-05  
 ; PRIOR APPLICATION NUMBER: 60/310,951  
 ; PRIOR FILING DATE: 2001-08-08  
 ; PRIOR APPLICATION NUMBER: 60/361,832  
 ; PRIOR FILING DATE: 2002-03-05  
 ; PRIOR APPLICATION NUMBER: 60/311,292  
 ; PRIOR FILING DATE: 2001-08-09  
 ; PRIOR APPLICATION NUMBER: 60/311,979  
 ; PRIOR FILING DATE: 2001-08-13  
 ; PRIOR APPLICATION NUMBER: 60/312,203

; PRIOR FILING DATE: 2001-08-14  
 ; PRIOR APPLICATION NUMBER: 60/313,201  
 ; PRIOR FILING DATE: 2001-08-17  
 ; PRIOR APPLICATION NUMBER: 60/313,702  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 191  
 ; SOFTWARE: Curaseq1 version 0.1  
 ; SEQ ID NO 34  
 ; LENGTH: 583  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-210-281-34

Query Match 0.9%; Score 9; DB 15; Length 583;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSES 952  
 Db 18 SEKSSSES 26

RESULT 14  
 US-10-087-192-549  
 ; Sequence 549, Application US/10087192  
 ; Publication No. US20020182586A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Morris, David W.  
 ; APPLICANT: Engelhard, Eric K.  
 ; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
 ; TITLE OF INVENTION: CANCER  
 ; FILE REFERENCE: 529452000122  
 ; CURRENT APPLICATION NUMBER: US/10/087,192  
 ; CURRENT FILING DATE: 2002-03-01  
 ; PRIOR APPLICATION NUMBER: US 09/747,377  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: US 09/798,586  
 ; PRIOR FILING DATE: 2001-03-02  
 ; NUMBER OF SEQ ID NOS: 2059  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 549  
 ; LENGTH: 608  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-10-087-192-549

Query Match 0.9%; Score 9; DB 13; Length 608;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSES 952  
 Db 36 SEKSSSES 44

RESULT 15  
 US-10-295-027-484  
 ; Sequence 484, Application US/10295027  
 ; Publication No. US20030232350A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Afar, Daniel  
 ; APPLICANT: Aziz, Natasha  
 ; APPLICANT: Ginsberg, Wendy M.  
 ; APPLICANT: Gish, Kurt C.  
 ; APPLICANT: Glynn, Richard  
 ; APPLICANT: Hevezi, Peter A.  
 ; APPLICANT: Mack, David H.  
 ; APPLICANT: Murray, Richard  
 ; APPLICANT: Watson, Susan R.  
 ; APPLICANT: Eos Biotechnology, Inc.  
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
 ; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer

```

; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 484
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-484

```

```

Query Match          0.9%; Score 9; DB 15; Length 883;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

```

```

RESULT 16
US-10-295-027-1330
; Sequence 1330, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Nataasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14

```

```

; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1330
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1330

```

```

Query Match          0.9%; Score 9; DB 15; Length 883;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

```

```

RESULT 17
US-10-173-999-28
; Sequence 28, Application US/10173999
; Publication No. US20040005563A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian
; FILE REFERENCE: 018501-002420US
; CURRENT APPLICATION NUMBER: US/10/173,999
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/299,234
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/315,287
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-173-999-28

```

```

Query Match          0.9%; Score 9; DB 15; Length 883;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 592 LTEIPVQAF 600
Db 174 LTEIPVQAF 182

```

```

RESULT 18
US-09-965-536A-16
; Sequence 16, Application US/09965536A
; Publication No. US20030027323A1
; GENERAL INFORMATION:
; APPLICANT: FEDER, J. N.
; APPLICANT: MINTIER, G.
; APPLICANT: RAMANATHAN, C. S.

```

; APPLICANT: HAWKEN, D. R.  
; TITLE OF INVENTION: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPBMY5,  
; FILE REFERENCE: D0041NP  
; CURRENT APPLICATION NUMBER: US/09/965,536A  
; CURRENT FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: 60/235,713  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/261,781  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: 60/306,605  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/310,436  
; PRIOR FILING DATE: 2001-08-03  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 907  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-09-965-536A-16

Query Match 0.9%; Score 9; DB 10; Length 907;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
|||  
DB 174 LTEIPVQAF 182

RESULT 19  
US-09-970-944-26  
; Sequence 26, Application US/09970944  
; Publication No. US20030204052A1  
; GENERAL INFORMATION:  
; APPLICANT: Herrman, John L  
; APPLICANT: Rastelli, Luca  
; TITLE OF INVENTION: No. US20030204052A1 Proteins and Nucleic Acids Encoding Same an  
; TITLE OF INVENTION: Antibodies Directed Against these Proteins  
; FILE REFERENCE: 21402-138  
; CURRENT APPLICATION NUMBER: US/09/970,944  
; CURRENT FILING DATE: 2002-05-02  
; PRIOR APPLICATION NUMBER: 60/237,862  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 907  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-970-944-26

Query Match 0.9%; Score 9; DB 10; Length 907;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
|||  
DB 174 LTEIPVQAF 182

RESULT 20  
US-09-970-944-27  
; Sequence 27, Application US/09970944  
; Publication No. US20030204052A1  
; GENERAL INFORMATION:  
; APPLICANT: Herrman, John L  
; APPLICANT: Rastelli, Luca  
; TITLE OF INVENTION: No. US20030204052A1 Proteins and Nucleic Acids Encoding Same an  
; TITLE OF INVENTION: Antibodies Directed Against these Proteins

; FILE REFERENCE: 21402-138  
; CURRENT APPLICATION NUMBER: US/09/970,944  
; CURRENT FILING DATE: 2002-05-02  
; PRIOR APPLICATION NUMBER: 60/237,862  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 907  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-970-944-27

Query Match 0.9%; Score 9; DB 10; Length 907;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
|||  
DB 174 LTEIPVQAF 182

RESULT 21  
US-10-251-385-264  
; Sequence 264, Application US/10251385  
; Publication No. US20030105292A1  
; GENERAL INFORMATION:  
; APPLICANT: Behan, Dominic P.  
; APPLICANT: Chalmers, Derek T.  
; APPLICANT: Liaw, Chen W.  
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G  
; TITLE OF INVENTION: Protein-Coupled  
; TITLE OF INVENTION: Receptors  
; FILE REFERENCE: AREN-0040  
; CURRENT APPLICATION NUMBER: US/10/251,385  
; CURRENT FILING DATE: 2002-09-20  
; PRIOR APPLICATION NUMBER: US/09/170,496  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 294  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 264  
; LENGTH: 907  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-251-385-264

Query Match 0.9%; Score 9; DB 14; Length 907;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
|||  
DB 174 LTEIPVQAF 182

RESULT 22  
US-10-251-385-278  
; Sequence 278, Application US/10251385  
; Publication No. US20030105292A1  
; GENERAL INFORMATION:  
; APPLICANT: Behan, Dominic P.  
; APPLICANT: Chalmers, Derek T.  
; APPLICANT: Liaw, Chen W.  
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G  
; TITLE OF INVENTION: Protein-Coupled  
; TITLE OF INVENTION: Receptors  
; FILE REFERENCE: AREN-0040  
; CURRENT APPLICATION NUMBER: US/10/251,385  
; CURRENT FILING DATE: 2002-09-20  
; PRIOR APPLICATION NUMBER: US/09/170,496  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 294  
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 278  
; LENGTH: 907  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-251-385-278

Query Match 0.9%; Score 9; DB 14; Length 907;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
|||  
Db 174 LTEIPVQAF 182

## RESULT 23

US-10-225-567A-422  
; Sequence 422, Application US/10225567A  
; Publication No. US2003013798A1

; GENERAL INFORMATION:

; APPLICANT: LifeSpan Biosciences

; APPLICANT: Brown, Joseph P.

; APPLICANT: Burner, Glenna C.

; APPLICANT: Roueh, Christine L.

; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS

; FILE REFERENCE: 1920-4-4

; CURRENT APPLICATION NUMBER: US/10/225,567A

; CURRENT FILING DATE: 2001-12-19

; PRIOR APPLICATION NUMBER: 60/257,144

; PRIOR FILING DATE: 2000-12-19

; NUMBER OF SEQ ID NOS: 2292

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 422

; LENGTH: 907

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-225-567A-422

Query Match 0.9%; Score 9; DB 14; Length 907;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
|||  
Db 174 LTEIPVQAF 182

## RESULT 24

US-10-295-027-849

; Sequence 849, Application US/10295027

; Publication No. US20030232350A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsberg, Wendy M.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Glynn, Richard

; APPLICANT: Hevezi, Peter A.

; APPLICANT: Mack, David H.

; APPLICANT: Murray, Richard

; APPLICANT: Watson, Susan R.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027

; CURRENT FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; PRIOR APPLICATION NUMBER: US 60/350,666

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/335,394

; PRIOR FILING DATE: 2001-11-15

; PRIOR APPLICATION NUMBER: US 60/332,464  
; PRIOR FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US 60/334,393  
; PRIOR FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/340,376  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: US 60/347,211  
; PRIOR FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: US 60/347,349  
; PRIOR FILING DATE: 2002-01-10  
; PRIOR APPLICATION NUMBER: US 60/355,250  
; PRIOR FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US 60/356,714  
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 1386

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 849

; LENGTH: 907

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-295-027-849

Query Match 0.9%; Score 9; DB 15; Length 907;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
|||  
Db 174 LTEIPVQAF 182

US-10-295-027-946

; Sequence 946, Application US/10295027

; Publication No. US20030232350A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsberg, Wendy M.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Glynn, Richard

; APPLICANT: Hevezi, Peter A.

; APPLICANT: Mack, David H.

; APPLICANT: Murray, Richard

; APPLICANT: Watson, Susan R.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027

; CURRENT FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; PRIOR APPLICATION NUMBER: US 60/350,666

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/335,394

; PRIOR FILING DATE: 2001-11-15

; PRIOR APPLICATION NUMBER: US 60/332,464

; PRIOR FILING DATE: 2001-11-21

; PRIOR APPLICATION NUMBER: US 60/334,393

; PRIOR FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: US 60/340,376

; PRIOR FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: US 60/347,211

; PRIOR FILING DATE: 2002-01-08

; PRIOR APPLICATION NUMBER: US 60/347,349

; PRIOR FILING DATE: 2002-01-10

; PRIOR APPLICATION NUMBER: US 60/355,250

; PRIOR FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: US 60/356,714

; PRIOR FILING DATE: 2002-02-13

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 1386  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 946  
; LENGTH: 907  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-295-027-946

Query Match 0.9%; Score 9; DB 15; Length 907;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600  
Db 174 LTEIPVQAF 182

## RESULT 26

US-10-295-027-1331

; Sequence 1331, Application US/10295027

; Publication No. US20030232350A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsberg, Wendy M.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Glynn, Richard

; APPLICANT: Hevezi, Peter A.

; APPLICANT: Mack, David H.

; APPLICANT: Murray, Richard

; APPLICANT: Watson, Susan R.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of diagnosis of Cancer, Compositions and

; FILE REFERENCE: Methods of Screening for Modulators of Cancer

; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027

; CURRENT FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; PRIOR APPLICATION NUMBER: US 60/350,666

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/335,394

; PRIOR FILING DATE: 2001-11-15

; PRIOR APPLICATION NUMBER: US 60/332,464

; PRIOR FILING DATE: 2001-11-21

; PRIOR APPLICATION NUMBER: US 60/334,393

; PRIOR FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: US 60/340,376

; PRIOR FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: US 60/347,211

; PRIOR FILING DATE: 2002-01-08

; PRIOR APPLICATION NUMBER: US 60/347,349

; PRIOR FILING DATE: 2002-01-10

; PRIOR APPLICATION NUMBER: US 60/355,250

; PRIOR FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: US 60/356,714

; PRIOR FILING DATE: 2002-02-13

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 1386

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1331

; LENGTH: 907

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-295-027-1331

Query Match

Best Local Similarity 0.9%; Score 9; DB 15; Length 907;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600  
Db 174 LTEIPVQAF 182

## RESULT 27

US-10-751-736-84

; Sequence 84, Application US/10751736

; Publication No. US20040265230A1

; GENERAL INFORMATION:

; APPLICANT: Wyeth

; APPLICANT: Martinez, Robert

; APPLICANT: Brown, Eugene

; APPLICANT: Liu, Wei

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING AND TREATING COLON

; FILE REFERENCE: AM100927 (031896-002000)

; CURRENT APPLICATION NUMBER: US/10/751,736

; CURRENT FILING DATE: 2003-01-06

; PRIOR APPLICATION NUMBER: US Provisional Application 60/438,000

; PRIOR FILING DATE: 2003-01-06

; NUMBER OF SEQ ID NOS: 54873

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 84

; LENGTH: 907

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-751-736-84

Query Match

Best Local Similarity 0.9%; Score 9; DB 16; Length 907;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600

Db 174 LTEIPVQAF 182

## RESULT 28

US-10-482-029-158

; Sequence 158, Application US/10482029

; Publication No. US20050037445A1

; GENERAL INFORMATION:

; APPLICANT: ODIN medical A/S

; TITLE OF INVENTION: Oncology drug innovation

; FILE REFERENCE: P 573 PC00

; CURRENT APPLICATION NUMBER: US/10/482,029

; CURRENT FILING DATE: 2003-12-29

; NUMBER OF SEQ ID NOS: 437

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 158

; LENGTH: 907

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-482-029-158

Query Match

Best Local Similarity 0.9%; Score 9; DB 17; Length 907;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 592 LTEIPVQAF 600

Db 174 LTEIPVQAF 182

## RESULT 29

US-10-505-486-103

; Sequence 103, Application US/10505486

; Publication No. US20050118639A1

; GENERAL INFORMATION:

; APPLICANT: Takeda Chemical Industries, Ltd.

; TITLE OF INVENTION: Determination of a ligand

; FILE REFERENCE: P03-0006PCT

; CURRENT APPLICATION NUMBER: US/10/505,486

; CURRENT FILING DATE: 2004-08-20

; PRIOR APPLICATION NUMBER: JP 2002-45728

; PRIOR FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: JP 2002-213949  
; PRIOR FILING DATE: 2002-07-23  
; PRIOR APPLICATION NUMBER: JP 2002-298237  
; PRIOR FILING DATE: 2002-10-11  
; NUMBER OF SEQ ID NOS: 233  
; SEQ ID NO 103  
; LENGTH: 1145  
; TYPE: PRT  
; ORGANISM: Human  
US-10-505-486-103

Query Match 0.9%; Score 9; DB 17; Length 1145;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVQAF 600  
Db 174 LTEIPVQAF 182

## RESULT 30

US-10-424-599-178222  
; Sequence 178222, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 178222  
; LENGTH: 56  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_131950C.1.pap  
US-10-424-599-178222

Query Match 0.8%; Score 8; DB 15; Length 56;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 645 DDNDGDEQ 652  
Db 4 DDNDGDEQ 11

## RESULT 31

US-10-424-599-245882  
; Sequence 245882, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 245882  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:

; NAME/KEY: unsure  
; LOCATION: (1)..(82)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_64062C.1.pap  
US-10-424-599-245882

Query Match 0.8%; Score 8; DB 15; Length 82;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 TGKLSLPR 301  
Db 8 TGKLSLPR 15

## RESULT 32

US-10-425-115-263516  
; Sequence 263516, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 263516  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_17193C.1.pap  
US-10-425-115-263516

Query Match 0.8%; Score 8; DB 16; Length 131;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSE 951  
Db 5 SEKSSSSE 12

## RESULT 33

US-10-425-115-306321  
; Sequence 306321, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 306321  
; LENGTH: 142  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(142)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_42429C.1.pap



US-10-425-115-306321

Query Match 0.8%; Score 8; DB 16; Length 142;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 EKIKDNLA 88  
|||||  
Db 99 EKIKDNLA 106  
|||||

RESULT 34

US-10-425-115-331445  
; Sequence 331445, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 331445  
; LENGTH: 179  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_65381C.1.ppp  
US-10-425-115-331445

Query Match 0.8%; Score 8; DB 16; Length 179;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 59 NETSGPVD 66  
|||||  
Db 139 NETSGPVD 146  
|||||

RESULT 35

US-10-424-599-280877  
; Sequence 280877, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J  
; APPLICANT: Kovalic, David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 280877  
; LENGTH: 183  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(183)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_95654C.1.ppp  
US-10-424-599-280877

Query Match 0.8%; Score 8; DB 15; Length 183;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 564 MGSKVKTL 571  
|||||  
Db 1 MGSKVKTL 8  
|||||

RESULT 36

US-10-059-911-12  
; Sequence 12, Application US/10059911  
; Publication No. US20030159180A1  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Mizukami, Yukiko  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants  
; FILE REFERENCE: 023070-090730US  
; CURRENT APPLICATION NUMBER: US/10/059,911  
; CURRENT FILING DATE: 2002-06-17  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 196  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:structurally  
; OTHER INFORMATION: altered ANT protein AWTIDNA  
US-10-059-911-12

Query Match 0.8%; Score 8; DB 14; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LKTVALTL 12  
|||||  
Db 177 LKTVALTL 184  
|||||

RESULT 37

US-10-732-923-5727  
; Sequence 5727, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 5727  
; LENGTH: 237  
; TYPE: PRT  
; ORGANISM: Nicotiana tabacum  
US-10-732-923-5727

Query Match 0.8%; Score 8; DB 17; Length 237;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 948 SSESANS 955  
|||||  
Db 42 SSESANS 49  
|||||

RESULT 38

US-10-425-115-262882  
; Sequence 262882, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 262882  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_17135C.1.pcp  
US-10-425-115-262882

Query Match 0.8%; Score 8; DB 16; Length 239;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 496 EASDDLEE 503  
|||||  
Db 1 EASDDLEE 8

## RESULT 39

US-10-059-911-24  
; Sequence 24, Application US/10059911  
; Publication No. US20030159180A1  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Mizukami, Yukiko  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants  
; FILE REFERENCE: 023070-090730US  
; CURRENT APPLICATION NUMBER: US/10/059,911  
; CURRENT FILING DATE: 2002-06-17  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: novel chimeric  
; OTHER INFORMATION: ANT protein ANT-AP25w2  
US-10-059-911-24

Query Match 0.8%; Score 8; DB 14; Length 264;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTTL 12  
|||||  
Db 245 LKTVALTTL 252

## RESULT 40

US-10-059-911-25  
; Sequence 25, Application US/10059911  
; Publication No. US20030159180A1  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Mizukami, Yukiko  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants  
; FILE REFERENCE: 023070-090730US  
; CURRENT APPLICATION NUMBER: US/10/059,911  
; CURRENT FILING DATE: 2002-06-17  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 25  
; LENGTH: 275

; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: novel chimeric  
; OTHER INFORMATION: ANT protein ANTdf1  
US-10-059-911-25

Query Match 0.8%; Score 8; DB 14; Length 275;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTTL 12  
|||||  
Db 256 LKTVALTTL 263

## RESULT 41

US-10-059-911-11  
; Sequence 11, Application US/10059911  
; Publication No. US20030159180A1  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Mizukami, Yukiko  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants  
; FILE REFERENCE: 023070-090730US  
; CURRENT APPLICATION NUMBER: US/10/059,911  
; CURRENT FILING DATE: 2002-06-17  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 277  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: structurally  
; OTHER INFORMATION: altered ANT protein ANTfDN3  
US-10-059-911-11

Query Match 0.8%; Score 8; DB 14; Length 277;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTTL 12  
|||||  
Db 258 LKTVALTTL 265

## RESULT 42

US-10-059-911-10  
; Sequence 10, Application US/10059911  
; Publication No. US20030159180A1  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Mizukami, Yukiko  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants  
; FILE REFERENCE: 023070-090730US  
; CURRENT APPLICATION NUMBER: US/10/059,911  
; CURRENT FILING DATE: 2002-06-17  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 308  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: structurally  
; OTHER INFORMATION: altered ANT protein ANTfDN2  
US-10-059-911-10

Query Match 0.8%; Score 8; DB 14; Length 308;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVLTL 12  
| | | | |  
Db 289 LKTVLTL 296

## RESULT 43

US-10-282-122A-45121  
; Sequence 45121, Application US/10282122A  
; Publication No. US20040029129A1

; GENERAL INFORMATION:  
; APPLICANT: Wang, Lianguo

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIORITY FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-03-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 45121

; LENGTH: 356

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-10-282-122A-45121

Query Match 0.8%; Score 8; DB 15; Length 356;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 997 LSLITAIK 1004  
| | | | |  
Db 156 LSLITAIK 163

## RESULT 44

US-10-032-585-7284

; Sequence 7284, Application US/10032585

; Publication No. US20030180953A1

; GENERAL INFORMATION:

; APPLICANT: Terry, Roemer D.

; APPLICANT: Bo, Jiang

; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/032,585  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7284  
; LENGTH: 411  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-10-032-585-7284

Query Match 0.8%; Score 8; DB 14; Length 411;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 LEKELDLL 802  
| | | | |  
Db 37 LEKELDLL 44

## RESULT 45

US-10-369-493-17401

; Sequence 17401, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 17401

; LENGTH: 458

; TYPE: PRT

; ORGANISM: Bacillus halodurans

US-10-369-493-17401

Query Match 0.8%; Score 8; DB 15; Length 458;

Best Local Similarity 100.0%; Pred. No. 2.6e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 KAIKAEK 757  
| | | | |  
Db 447 KAIKAEK 454

## RESULT 46

US-10-059-911-9

; Sequence 9, Application US/10059911

; Publication No. US20030159180A1

; GENERAL INFORMATION:

; APPLICANT: Fischer, Robert L.

; APPLICANT: Mizukami, Yukiko

; TITLE OF INVENTION: The Regents of the University of California

; FILE REFERENCE: 023070-090730US

; CURRENT APPLICATION NUMBER: US/10/059,911

; CURRENT FILING DATE: 2002-06-17

; NUMBER OF SEQ ID NOS: 42

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 9

; LENGTH: 473

; TYPE: PRT

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:structurally
; OTHER INFORMATION: altered ANT protein ANTND1
US-10-059-911-9
Query Match      0.8%; Score 8; DB 14; Length 473;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 LKTVALTL 12
Db      454 LKTVALTL 461

RESULT 47
US-10-481-032A-532
; Sequence 532, Application US/10481032A
; Publication No. US20050177901A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Cheng, Wengqiong
; APPLICANT: Briggs, Steven
; APPLICANT: Cooper, Bret
; APPLICANT: Goff, Stephen A.
; APPLICANT: Moughamer, Todd
; APPLICANT: Glazebrook, Jane
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreppe, Joel
; APPLICANT: Provart, Nicolas
; APPLICANT: Rickes, Darrell
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 60148USPCT
; CURRENT APPLICATION NUMBER: US/10/481,032A
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: US 60/300,112
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/342,327
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: PCT/IB02/02450
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 1201
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 532
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-481-032A-532

Query Match      0.8%; Score 8; DB 18; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 48
US-10-389-566-712
; Sequence 712, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
```

```
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 712
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-389-566-712

Query Match      0.8%; Score 8; DB 15; Length 488;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 49
US-10-389-566-731
; Sequence 731, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 731
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-389-566-731

Query Match      0.8%; Score 8; DB 15; Length 488;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 50
US-10-389-566-1366
; Sequence 1366, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
```

```
; SOFTWARE: PatehtIn version 3.2
; SEQ ID NO 1366
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-389-566-1366

Query Match          0.8%; Score 8; DB 15; Length 488;
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 51
US-10-389-566-1367
; Sequence 1367, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1367
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-389-566-1367

Query Match          0.8%; Score 8; DB 15; Length 488;
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 52
US-10-481-032A-104
; Sequence 104, Application US/10481032A
; Publication No. US20050177901A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Cheng, Wengdong
; APPLICANT: Briggs, Steven
; APPLICANT: Cooper, Bret
; APPLICANT: Goff, Stephen A.
; APPLICANT: Moughamer, Todd
; APPLICANT: Glazebrook, Jane
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicolas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 60148USPCT
; CURRENT APPLICATION NUMBER: US/10/481,032A
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: US 60/300,112
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/325,277
```

```
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/342,327
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: PCT/IB02/02450
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 1201
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 104
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Oryza sativa
; NAME/KEY: misc feature
; LOCATION: (286)..(286)
; OTHER INFORMATION: Xaa = any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (300)..(300)
; OTHER INFORMATION: Xaa = any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (451)..(451)
; OTHER INFORMATION: Xaa = any naturally occurring amino acid
US-10-481-032A-104

Query Match          0.8%; Score 8; DB 18; Length 496;
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 53
US-10-389-566-2286
; Sequence 2286, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2286
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-10-389-566-2286

Query Match          0.8%; Score 8; DB 15; Length 500;
Best Local Similarity 100.0%; Pred. No. 2.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

Qy      804 GLVEGKGP 811
Db      51 GLVEGKGP 58

RESULT 54
US-10-389-566-860
; Sequence 860, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
```

```
/ APPLICANT: Monsanto Technology, LLC
/ APPLICANT: Laurie, Cathy C
/ TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
/ FILE REFERENCE: 38-77(52900)D
/ CURRENT APPLICATION NUMBER: US/10/389,566
/ CURRENT FILING DATE: 2003-03-31
/ PRIOR APPLICATION NUMBER: US 60/365,301
/ PRIOR FILING DATE: 2002-03-15
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/392,018
/ PRIOR FILING DATE: 2002-06-26
/ NUMBER OF SEQ ID NOS: 2459
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 860
/ LENGTH: 503
/ TYPE: PRT
/ ORGANISM: Triticum aestivum
/ US-10-389-566-860

Query Match      0.8%; Score 8; DB 15; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 55
US-10-389-566-1411
/ Sequence 1411, Application US/10389566
/ Publication No. US20040025202A1
/ GENERAL INFORMATION:
/ APPLICANT: Monsanto Technology, LLC
/ APPLICANT: Laurie, Cathy C
/ TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
/ FILE REFERENCE: 38-77(52900)D
/ CURRENT APPLICATION NUMBER: US/10/389,566
/ CURRENT FILING DATE: 2003-03-31
/ PRIOR APPLICATION NUMBER: US 60/365,301
/ PRIOR FILING DATE: 2002-03-15
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/392,018
/ PRIOR FILING DATE: 2002-06-26
/ NUMBER OF SEQ ID NOS: 2459
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 1411
/ LENGTH: 503
/ TYPE: PRT
/ ORGANISM: Triticum aestivum
/ US-10-389-566-1411

Query Match      0.8%; Score 8; DB 15; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 56
US-10-389-566-2188
/ Sequence 2188, Application US/10389566
/ Publication No. US20040025202A1
/ GENERAL INFORMATION:
/ APPLICANT: Monsanto Technology, LLC
/ APPLICANT: Laurie, Cathy C
/ TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
/ FILE REFERENCE: 38-77(52900)D
/ CURRENT APPLICATION NUMBER: US/10/389,566
```

```
/ CURRENT FILING DATE: 2003-03-31
/ PRIOR APPLICATION NUMBER: US 60/365,301
/ PRIOR FILING DATE: 2002-03-15
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/392,018
/ PRIOR FILING DATE: 2002-06-26
/ NUMBER OF SEQ ID NOS: 2459
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 2188
/ LENGTH: 503
/ TYPE: PRT
/ ORGANISM: Secale cereale
/ US-10-389-566-2188

Query Match      0.8%; Score 8; DB 15; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 57
US-10-389-566-859
/ Sequence 859, Application US/10389566
/ Publication No. US20040025202A1
/ GENERAL INFORMATION:
/ APPLICANT: Monsanto Technology, LLC
/ APPLICANT: Laurie, Cathy C
/ TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
/ FILE REFERENCE: 38-77(52900)D
/ CURRENT APPLICATION NUMBER: US/10/389,566
/ CURRENT FILING DATE: 2003-03-31
/ PRIOR APPLICATION NUMBER: US 60/365,301
/ PRIOR FILING DATE: 2002-03-15
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/392,018
/ PRIOR FILING DATE: 2002-06-26
/ NUMBER OF SEQ ID NOS: 2459
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 859
/ LENGTH: 527
/ TYPE: PRT
/ ORGANISM: Triticum aestivum
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (19)..(19)
/ OTHER INFORMATION: Xaa can be any naturally occurring amino acid
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (496)..(497)
/ OTHER INFORMATION: Xaa can be any naturally occurring amino acid
/ US-10-389-566-859

Query Match      0.8%; Score 8; DB 15; Length 527;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      804 GLVEGKGP 811
Db      57 GLVEGKGP 64

RESULT 58
US-10-389-566-1020
/ Sequence 1020, Application US/10389566
/ Publication No. US20040025202A1
/ GENERAL INFORMATION:
/ APPLICANT: Monsanto Technology, LLC
/ APPLICANT: Laurie, Cathy C
```

```
/ TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
/ FILE REFERENCE: 38-77(52900)D
/ CURRENT APPLICATION NUMBER: US/10/389,566
/ CURRENT FILING DATE: 2003-03-31
/ PRIOR APPLICATION NUMBER: US 60/365,301
/ PRIOR FILING DATE: 2002-03-15
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/392,018
/ PRIOR FILING DATE: 2002-06-26
/ NUMBER OF SEQ ID NOS: 2459
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 1020
/ LENGTH: 533
/ TYPE: PRT
/ ORGANISM: Hordeum vulgare
US-10-389-566-1020

Query Match          0.8%; Score 8; DB 15; Length 533;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      53 GLVEGKGP 60

RESULT 59
US-10-389-566-1021
/ Sequence 1021, Application US/10389566
/ Publication No. US20040025202A1
/ GENERAL INFORMATION:
/ APPLICANT: Monsanto Technology, LLC
/ APPLICANT: Laurie, Cathy C
/ TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
/ FILE REFERENCE: 38-77(52900)D
/ CURRENT APPLICATION NUMBER: US/10/389,566
/ CURRENT FILING DATE: 2003-03-31
/ PRIOR APPLICATION NUMBER: US 60/365,301
/ PRIOR FILING DATE: 2002-03-15
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/392,018
/ PRIOR FILING DATE: 2002-06-26
/ NUMBER OF SEQ ID NOS: 2459
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 1021
/ LENGTH: 533
/ TYPE: PRT
/ ORGANISM: Hordeum vulgare
US-10-389-566-1021

Query Match          0.8%; Score 8; DB 15; Length 533;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      53 GLVEGKGP 60

RESULT 60
US-10-389-566-1041
/ Sequence 1041, Application US/10389566
/ Publication No. US20040025202A1
/ GENERAL INFORMATION:
/ APPLICANT: Monsanto Technology, LLC
/ APPLICANT: Laurie, Cathy C
/ TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
/ FILE REFERENCE: 38-77(52900)D
/ CURRENT APPLICATION NUMBER: US/10/389,566
/ CURRENT FILING DATE: 2003-03-31
/ PRIOR APPLICATION NUMBER: US 60/365,301
```

```
/ PRIOR FILING DATE: 2002-03-15
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/392,018
/ PRIOR FILING DATE: 2002-06-26
/ NUMBER OF SEQ ID NOS: 2459
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 1041
/ LENGTH: 535
/ TYPE: PRT
/ ORGANISM: Hordeum vulgare
US-10-389-566-1041

Query Match          0.8%; Score 8; DB 15; Length 535;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 61
US-10-389-566-1124
/ Sequence 1124, Application US/10389566
/ Publication No. US20040025202A1
/ GENERAL INFORMATION:
/ APPLICANT: Monsanto Technology, LLC
/ APPLICANT: Laurie, Cathy C
/ TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
/ FILE REFERENCE: 38-77(52900)D
/ CURRENT APPLICATION NUMBER: US/10/389,566
/ CURRENT FILING DATE: 2003-03-31
/ PRIOR APPLICATION NUMBER: US 60/365,301
/ PRIOR FILING DATE: 2002-03-15
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/392,018
/ PRIOR FILING DATE: 2002-06-26
/ NUMBER OF SEQ ID NOS: 2459
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 1124
/ LENGTH: 535
/ TYPE: PRT
/ ORGANISM: Hordeum vulgare
US-10-389-566-1124

Query Match          0.8%; Score 8; DB 15; Length 535;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 62
US-10-389-566-2123
/ Sequence 2123, Application US/10389566
/ Publication No. US20040025202A1
/ GENERAL INFORMATION:
/ APPLICANT: Monsanto Technology, LLC
/ APPLICANT: Laurie, Cathy C
/ TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
/ FILE REFERENCE: 38-77(52900)D
/ CURRENT APPLICATION NUMBER: US/10/389,566
/ CURRENT FILING DATE: 2003-03-31
/ PRIOR APPLICATION NUMBER: US 60/365,301
/ PRIOR FILING DATE: 2002-03-15
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/392,018
/ PRIOR FILING DATE: 2002-06-26
```



```
/ NUMBER OF SEQ ID NOS: 2459
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 2123
/ LENGTH: 535
/ TYPE: PRT
/ ORGANISM: Hordeum vulgare
US-10-389-566-2123

Query Match      0.8%; Score 8; DB 15; Length 535;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 63
US-10-389-566-2186
/ Sequence 2186, Application US/10389566
/ Publication No. US20040025202A1
/ GENERAL INFORMATION:
/ APPLICANT: Monsanto Technology, LLC
/ APPLICANT: Laurie, Cathy C
/ TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
/ FILE REFERENCE: 38-77(52900)D
/ CURRENT APPLICATION NUMBER: US/10/389,566
/ CURRENT FILING DATE: 2003-03-31
/ PRIOR APPLICATION NUMBER: US 60/365,301
/ PRIOR FILING DATE: 2002-03-15
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/392,018
/ PRIOR FILING DATE: 2002-06-26
/ NUMBER OF SEQ ID NOS: 2459
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 2186
/ LENGTH: 535
/ TYPE: PRT
/ ORGANISM: Hordeum vulgare
US-10-389-566-2186

Query Match      0.8%; Score 8; DB 15; Length 535;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 64
US-10-389-566-2397
/ Sequence 2397, Application US/10389566
/ Publication No. US20040025202A1
/ GENERAL INFORMATION:
/ APPLICANT: Monsanto Technology, LLC
/ APPLICANT: Laurie, Cathy C
/ TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
/ FILE REFERENCE: 38-77(52900)D
/ CURRENT APPLICATION NUMBER: US/10/389,566
/ CURRENT FILING DATE: 2003-03-31
/ PRIOR APPLICATION NUMBER: US 60/365,301
/ PRIOR FILING DATE: 2002-03-15
/ PRIOR APPLICATION NUMBER: US 60/391,786
/ PRIOR FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/392,018
/ PRIOR FILING DATE: 2002-06-26
/ NUMBER OF SEQ ID NOS: 2459
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 2397
/ LENGTH: 535
/ TYPE: PRT

/ ORGANISM: Hordeum vulgare
US-10-389-566-2397

Query Match      0.8%; Score 8; DB 15; Length 535;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      804 GLVEGKGP 811
Db      55 GLVEGKGP 62

RESULT 65
US-10-059-911-2
/ Sequence 2, Application US/10059911
/ Publication No. US20030159180A1
/ GENERAL INFORMATION:
/ APPLICANT: Fischer, Robert L.
/ APPLICANT: Mizukami, Yukiko
/ APPLICANT: The Regents of the University of California
/ TITLE OF INVENTION: Methods for Altering Organ Mass in Plants
/ FILE REFERENCE: 023070-090730US
/ CURRENT APPLICATION NUMBER: US/10/059,911
/ CURRENT FILING DATE: 2002-06-17
/ NUMBER OF SEQ ID NOS: 42
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 555
/ TYPE: PRT
/ ORGANISM: Arabidopsis thaliana
/ FEATURE:
/ OTHER INFORMATION: Arabidopsis AINTEGUMENTA (ANT)
US-10-059-911-2

Query Match      0.8%; Score 8; DB 14; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 LKTVALTL 12
Db      536 LKTVALTL 543

RESULT 66
US-10-059-911-21
/ Sequence 21, Application US/10059911
/ Publication No. US20030159180A1
/ GENERAL INFORMATION:
/ APPLICANT: Fischer, Robert L.
/ APPLICANT: Mizukami, Yukiko
/ APPLICANT: The Regents of the University of California
/ TITLE OF INVENTION: Methods for Altering Organ Mass in Plants
/ FILE REFERENCE: 023070-090730US
/ CURRENT APPLICATION NUMBER: US/10/059,911
/ CURRENT FILING DATE: 2002-06-17
/ NUMBER OF SEQ ID NOS: 42
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 21
/ LENGTH: 555
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:structurally
/ OTHER INFORMATION: altered ANT protein ANTMNLS
US-10-059-911-21

Query Match      0.8%; Score 8; DB 14; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 LKTVALTL 12
Db      536 LKTVALTL 543
```

```

Best Local Similarity 100.0%; Pred. No. 3.1e+02; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

Qy 5 LKTVALTL 12
Db 536 LKTVALTL 543

RESULT 69
US-10-183-687-424
; Sequence 424, Application US/10183687
; Publication No. US20030204870A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Allen, William B.
; APPLICANT: Cahoon, Rebecca
; APPLICANT: Epelbaum, Sabine
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Harvell, Leslie T.
; APPLICANT: Jones, Todd
; APPLICANT: Kinney, Tony
; APPLICANT: Klein, Ted
; APPLICANT: Li, Changjiang
; APPLICANT: Oliveira, Igor Cunha
; APPLICANT: Sakai, Hajime
; APPLICANT: Shen, Bo
; APPLICANT: Tarczynski, Mitchell C.
; TITLE OF INVENTION: Alteration Of Oil Traits In Plants
; FILE REFERENCE: BB1458 US NA
; CURRENT APPLICATION NUMBER: US/10/183,687
; CURRENT FILING DATE: 2002-06-27
; PRIOR APPLICATION NUMBER: 60/301,913
; PRIOR FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 424
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: gi 1171429
US-10-183-687-424

Query Match 0.8%; Score 8; DB 15; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LKTVALTL 12
Db 536 LKTVALTL 543

RESULT 70
US-10-623-477-2
; Sequence 2, Application US/10623477
; Publication No. US2005013245A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Methods for Altering Organ Mass, Controlling Fertility
; FILE REFERENCE: 023070-090720US
; CURRENT APPLICATION NUMBER: US/10/623,477
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US/09/479,855
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 555

```

---

```

US-10-059-911-22
; Sequence 22, Application US/10059911
; Publication No. US20030159180A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Mizukami, Yukiko
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Methods for Altering Organ Mass in Plants
; FILE REFERENCE: 023070-090730US
; CURRENT APPLICATION NUMBER: US/10/059,911
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:structurally
; OTHER INFORMATION: altered ANT protein ANTMRII, temperature sensitive
; OTHER INFORMATION: mutant
US-10-059-911-22

Query Match 0.8%; Score 8; DB 14; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LKTVALTL 12
Db 536 LKTVALTL 543

RESULT 68
US-10-183-687-422
; Sequence 422, Application US/10183687
; Publication No. US20030204870A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Allen, William B.
; APPLICANT: Cahoon, Rebecca
; APPLICANT: Epelbaum, Sabine
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Harvell, Leslie T.
; APPLICANT: Jones, Todd
; APPLICANT: Kinney, Tony
; APPLICANT: Klein, Ted
; APPLICANT: Li, Changjiang
; APPLICANT: Oliveira, Igor Cunha
; APPLICANT: Sakai, Hajime
; APPLICANT: Shen, Bo
; APPLICANT: Tarczynski, Mitchell C.
; TITLE OF INVENTION: Alteration Of Oil Traits In Plants
; FILE REFERENCE: BB1458 US NA
; CURRENT APPLICATION NUMBER: US/10/183,687
; CURRENT FILING DATE: 2002-06-27
; PRIOR APPLICATION NUMBER: 60/301,913
; PRIOR FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 422
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: gi 2129537
US-10-183-687-422

Query Match 0.8%; Score 8; DB 15; Length 555;

```

;  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: AINTEGUMENTA (ANT)  
US-10-623-477-2

Query Match 0.8%; Score 8; DB 18; Length 555;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVLTL 12  
| | | | |  
Db 536 LKTVLTL 543

## RESULT 71

US-10-739-930-6762  
; Sequence 6762, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; FILE REFERENCE: 38-21(53377)B  
; CURRENT APPLICATION NUMBER: US/10/739,930  
; CURRENT FILING DATE: 2003-12-18  
; NUMBER OF SEQ ID NOS: 11088  
; SEQ ID NO 6762  
; LENGTH: 642  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C8623\_1.p  
US-10-739-930-6762

Query Match 0.8%; Score 8; DB 16; Length 642;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LTVSVVT 19  
| | | | |  
Db 149 LTVSVVT 156

## RESULT 72

US-09-864-761-47959  
; Sequence 47959, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmics-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664

;  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 47959  
; LENGTH: 660  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AJ010770.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1  
; OTHER INFORMATION: SWISSPROT HIT: Q99323, EVALUE 3.00e-17  
; OTHER INFORMATION: EST\_HUMAN HIT: AU132932.1, EVALUE 1.00e-105  
US-09-864-761-47959

Query Match 0.8%; Score 8; DB 9; Length 660;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKDKL 523  
| | | | |  
Db 35 TLELKDKL 42

## RESULT 73

US-09-981-397A-22  
; Sequence 22, Application US/09981397A  
; Publication No. US20030082519A1  
; GENERAL INFORMATION:  
; APPLICANT: Axxima Pharmaceuticals AG  
; APPLICANT: Schubart, Daniel  
; APPLICANT: Habenberger, Peter  
; APPLICANT: Stein-Gerlach, Matthias  
; APPLICANT: Bevec, Dorian  
; TITLE OF INVENTION: Cellular Kinases Involved in Cytomegalovirus Infection and their  
; FILE REFERENCE: AXM-004.1 US  
; CURRENT APPLICATION NUMBER: US/09/981,397A  
; CURRENT FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: 60/240,750  
; PRIOR FILING DATE: 2000-10-16  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 22  
; LENGTH: 686  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-981-397A-22

Query Match 0.8%; Score 8; DB 10; Length 686;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSSE 951  
|||||  
Db 5 SEKSSSSE 12

## RESULT 74

US-10-756-149-4945  
; Sequence 4945, Application US/10756149  
; Publication No. US20050181375A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER  
; FILE REFERENCE: file  
; CURRENT APPLICATION NUMBER: US/10/756,149  
; CURRENT FILING DATE: 2004-01-12  
; NUMBER OF SEQ ID NOS: 5818  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4945  
; LENGTH: 686  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-756-149-4945

Query Match 0.8%; Score 8; DB 18; Length 686;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSSE 951  
|||||  
Db 5 SEKSSSSE 12

## RESULT 75

US-10-618-941-102  
; Sequence 102, Application US/10618941  
; Publication No. US2004019792A1  
; GENERAL INFORMATION:  
; APPLICANT: WHYTE, DAVID  
; APPLICANT: MANNING, GERARD  
; APPLICANT: CAENEPEEL, SEAN  
; TITLE OF INVENTION: NOVEL KINASES  
; FILE REFERENCE: 034536-0321  
; CURRENT APPLICATION NUMBER: US/10/618,941  
; CURRENT FILING DATE: 2003-07-15  
; PRIOR APPLICATION NUMBER: 60/395,632  
; PRIOR FILING DATE: 2002-07-15  
; NUMBER OF SEQ ID NOS: 143  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 102  
; LENGTH: 688  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-618-941-102

Query Match 0.8%; Score 8; DB 16; Length 688;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSSE 951  
|||||  
Db 5 SEKSSSSE 12

## RESULT 76

US-10-425-114-54382  
; Sequence 54382, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 54382  
; LENGTH: 715  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3063-134-E5\_FLI.pep  
US-10-425-114-54382

Query Match 0.8%; Score 8; DB 15; Length 715;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 944 SEKSSSSE 951  
|||||  
Db 32 SEKSSSSE 39

## RESULT 77

US-10-029-386-32324  
; Sequence 32324, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: AEOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 32324  
; LENGTH: 820  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC000065.1  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.87  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3  
; OTHER INFORMATION: SWISSPROT HIT: Q99996, EVALUUE 0.00e+00  
US-10-029-386-32324

Query Match 0.8%; Score 8; DB 14; Length 820;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKDKL 523  
|||||  
Db 51 TLELKDKL 58

## RESULT 78

US-10-732-923-13561  
; Sequence 13561, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C

```
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 13561
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Sycon raphanus
US-10-732-923-13561

Query Match      0.8%; Score 8; DB 17; Length 879;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      732 KOKFLOEA 739
Db      659 KOKFLOEA 666
      |||||

RESULT 79
US-10-473-574-22
; Sequence 22, Application US/10473574
; Publication No. US20040116670A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; HAPALIA, April J.A.;
; APPLICANT: TANG, Y. Tom; YUE, Henry;
; APPLICANT: KHAN, Farrah A.; ISON, Craig H.;
; APPLICANT: BAUGHN, Mariah R.; WARREN, Bridget A.;
; APPLICANT: DUGGAN, Brendan M.; THANGAVELU, Kavitha;
; APPLICANT: HONCHELL, Cynthia D.; AZIMZAI, Yalda;
; APPLICANT: ELLIOTT, Vicki S.; BURFORD, Neil;
; APPLICANT: DING, Li; YUE, Huibin;
; APPLICANT: BECHA, Shanya; EMERLING, Brooke M.;
; APPLICANT: RICHARDSON, Thomas W.; LEE, Soo Yeun;
; APPLICANT: BANDMAN, Olga; LAL, Preeti G.;
; APPLICANT: LEE, Sally; GIBTZEN, Kimberly J.;
; APPLICANT: CHAWLA, Narinder K.; GRIFFIN, Jennifer A.;
; APPLICANT: LEE, Ernestine A.; SWARNAKAR, Anita;
; APPLICANT: RING, Huijun Z.; JONES, Karen Anne
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0918 USN
; CURRENT APPLICATION NUMBER: US/10/473,574
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US02/09288
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: US 60/294,451
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 60/291,870
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/290,518
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/288,609
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/283,769
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/281,323
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/280,508
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 916
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7468507CD1
US-10-473-574-22

Query Match      0.8%; Score 8; DB 16; Length 916;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      793 KLEKELD 800
Db      707 KLEKELD 714
      |||||

RESULT 80
US-10-170-385-39
; Sequence 39, Application US/10170385
; Publication No. US20030203372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingsman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; FILE REFERENCE: 532682000100
; CURRENT APPLICATION NUMBER: US/10/170,385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 931
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-170-385-39

Query Match      0.8%; Score 8; DB 15; Length 931;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      945 EKSSSES 952
Db      892 EKSSSES 899
      |||||

RESULT 81
US-10-408-765A-1585
; Sequence 1585, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1585
; LENGTH: 931
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1585

Query Match      0.8%; Score 8; DB 16; Length 931;
```

```

Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 945 EKSSSES 952
Db 892 EKSSSES 899

RESULT 82
US-09-969-528-2
; Sequence 2, Application US/09969528
; Patent No. US20020150567A1
; GENERAL INFORMATION:
; APPLICANT: Pot, David A.
; Williams, Lewis T.
; Jefferson, Anne Bennett
; Majerus, Philip W.
; TITLE OF INVENTION: No. US20020150567A1el Grb2 Associating Protein and Nucleic
; Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/969,528
; FILING DATE: 01-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/560,005
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0624000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-969-528-2

Query Match 0.8%; Score 8; DB 9; Length 976;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 901 SKLSQLTS 908
Db 50 SKLSQLTS 57

RESULT 83
US-10-080-608A-11
; Sequence 11, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES

```

```
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-2

Query Match      0.8%; Score 8; DB 14; Length 3907;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      516 TLELKDKL 523
Db      345 TLELKDKL 352

RESULT 86
US-10-756-149-5739
; Sequence 5739, Application US/10756149
; Publication No. US20050181375A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756,149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5739
; LENGTH: 3908
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-756-149-5739

Query Match      0.8%; Score 8; DB 18; Length 3908;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      516 TLELKDKL 523
Db      345 TLELKDKL 352

RESULT 87
US-10-370-685-100
; Sequence 100, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; APPLICANT: Goldberg, Edward
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF P-004
; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
```

```
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 100
; LENGTH: 3911
; TYPE: PRT
; ORGANISM: human
US-10-370-685-100

Query Match      0.8%; Score 8; DB 15; Length 3911;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      516 TLELKDKL 523
Db      357 TLELKDKL 364

RESULT 88
US-10-408-765A-1839
; Sequence 1839, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1839
; LENGTH: 3911
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1839

Query Match      0.8%; Score 8; DB 16; Length 3911;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      516 TLELKDKL 523
Db      357 TLELKDKL 364

RESULT 89
US-10-171-311-8
; Sequence 8, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
```



```
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 3917
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-8
```

```
Query Match          0.8%; Score 8; DB 14; Length 3917;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 516 TLELKDKL 523
Db 345 TLELKDKL 352
```

```
RESULT 90
US-10-171-311-6
; Sequence 6, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoarsh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-6
```

```
Query Match          0.8%; Score 8; DB 14; Length 3925;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 516 TLELKDKL 523
Db 345 TLELKDKL 352
```

```
RESULT 91
US-10-024-918-8
; Sequence 8, Application US/10024918
; Publication No. US20020168718A1
; GENERAL INFORMATION:
; APPLICANT: Hubbell, Jeffrey
; APPLICANT: Schense, Jason
; APPLICANT: Zisch, Andreas
; APPLICANT: Hall, Heike
; TITLE OF INVENTION: ENZYME-MEDIATED MODIFICATION OF FIBRIN FOR TISSUE ENGINEERING
; FILE REFERENCE: CIT 2606 CIP
```

```
; CURRENT APPLICATION NUMBER: US/10/024,918
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 16
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: peptide with glutamine at the transglutaminase coupling site and
; OTHER INFORMATION: an active peptide that mimics the crosslinking site in the gamma
; NAME/KEY: MOD.RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: dansylated tyrosine
US-10-024-918-8
```

```
Query Match          0.7%; Score 7; DB 13; Length 16;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 855 DTIGEGQ 861
Db 4 DTIGEGQ 10
```

```
RESULT 92
US-10-024-918-12
; Sequence 12, Application US/10024918
; Publication No. US20020168718A1
; GENERAL INFORMATION:
; APPLICANT: Hubbell, Jeffrey
; APPLICANT: Schense, Jason
; APPLICANT: Zisch, Andreas
; APPLICANT: Hall, Heike
; TITLE OF INVENTION: ENZYME-MEDIATED MODIFICATION OF FIBRIN FOR TISSUE ENGINEERING
; FILE REFERENCE: CIT 2606 CIP
; CURRENT APPLICATION NUMBER: US/10/024,918
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 16
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: peptide with glutamine at the transglutaminase coupling site in t
; OTHER INFORMATION: he chain of fibrinogen
US-10-024-918-12
```

```
Query Match          0.7%; Score 7; DB 13; Length 16;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 855 DTIGEGQ 861
Db 4 DTIGEGQ 10
```

```
RESULT 93
US-10-325-021-13
; Sequence 13, Application US/10325021
; Publication No. US20030168833A1
; GENERAL INFORMATION:
; APPLICANT: Itolf, Matthias
; APPLICANT: Schense, Jason
; APPLICANT: Hubbell, Jeffery Alan
; APPLICANT: Jen, Anna
; TITLE OF INVENTION: Growth Factor Modified Protein Matrices for Tissue
; FILE REFERENCE: ETH 107 CIP(3)
; CURRENT APPLICATION NUMBER: US/10/325,021
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 10/024,918
```

; PRIOR FILING DATE: 2001-12-18  
 ; PRIOR APPLICATION NUMBER: PCT/EP02/12458  
 ; PRIOR FILING DATE: 2002-11-07  
 ; NUMBER OF SEQ ID NOS: 23  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 13  
 ; LENGTH: 16  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-325-021-13

Query Match 0.7%; Score 7; DB 14; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 855 DTIGEGQ 861  
 |||||  
 DB 4 DTIGEGQ 10

RESULT 94  
 US-10-650-509-8  
 ; Sequence 8, Application US/10650509  
 ; Publication No. US20040082513A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hubbell, Jeffrey  
 ; APPLICANT: Schense, Jason  
 ; APPLICANT: Zisch, Andreas  
 ; APPLICANT: Hall, Heike  
 ; TITLE OF INVENTION: ENZYME-MEDIATED MODIFICATION OF FIBRIN FOR TISSUE ENGINEERING  
 ; FILE REFERENCE: CIT 2606 CIP CON  
 ; CURRENT APPLICATION NUMBER: US/10/650,509  
 ; CURRENT FILING DATE: 2003-08-27  
 ; PRIOR APPLICATION NUMBER: 10/024,918  
 ; PRIOR FILING DATE: 2001-12-18  
 ; PRIOR APPLICATION NUMBER: 09/057,052  
 ; PRIOR FILING DATE: 1998-04-08  
 ; PRIOR APPLICATION NUMBER: PCT/US98/06617  
 ; PRIOR FILING DATE: 1998-04-02  
 ; PRIOR APPLICATION NUMBER: 60/042,143  
 ; PRIOR FILING DATE: 1997-04-03  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 8  
 ; LENGTH: 16  
 ; TYPE: PRT  
 ; ORGANISM: artificial sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: peptide with glutamine at the transglutaminase coupling site and  
 ; OTHER INFORMATION: an active peptide that mimics the crosslinking site in the gamma  
 ; OTHER INFORMATION: chain of fibrinogen  
 ; FEATURE:  
 ; NAME/KEY: MOD RES  
 ; LOCATION: (1)..(1)  
 ; OTHER INFORMATION: dansylated tyrosine  
 US-10-650-509-8

Query Match 0.7%; Score 7; DB 15; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 855 DTIGEGQ 861  
 |||||  
 DB 4 DTIGEGQ 10

RESULT 95  
 US-10-650-509-12  
 ; Sequence 12, Application US/10650509  
 ; Publication No. US20040082513A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hubbell, Jeffrey  
 ; APPLICANT: Schense, Jason

; APPLICANT: Zisch, Andreas  
 ; APPLICANT: Hall, Heike  
 ; TITLE OF INVENTION: ENZYME-MEDIATED MODIFICATION OF FIBRIN FOR TISSUE ENGINEERING  
 ; FILE REFERENCE: CIT 2606 CIP CON  
 ; CURRENT APPLICATION NUMBER: US/10/650,509  
 ; CURRENT FILING DATE: 2003-08-27  
 ; PRIOR APPLICATION NUMBER: 10/024,918  
 ; PRIOR FILING DATE: 2001-12-18  
 ; PRIOR APPLICATION NUMBER: 09/057,052  
 ; PRIOR FILING DATE: 1998-04-08  
 ; PRIOR APPLICATION NUMBER: PCT/US98/06617  
 ; PRIOR FILING DATE: 1998-04-02  
 ; PRIOR APPLICATION NUMBER: 60/042,143  
 ; PRIOR FILING DATE: 1997-04-03  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 12  
 ; LENGTH: 16  
 ; TYPE: PRT  
 ; ORGANISM: artificial sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: peptide with glutamine at the transglutaminase coupling site in  
 ; OTHER INFORMATION: he chain of fibrinogen  
 US-10-650-509-12

Query Match 0.7%; Score 7; DB 15; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 855 DTIGEGQ 861  
 |||||  
 DB 4 DTIGEGQ 10

RESULT 96  
 US-10-343-663A-36  
 ; Sequence 36, Application US/10343663A  
 ; Publication No. US20040106148A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dowler, S.  
 ; APPLICANT: Campbell, D.  
 ; APPLICANT: Gray A.  
 ; APPLICANT: Downes, P.  
 ; APPLICANT: Alessi, D.  
 ; TITLE OF INVENTION: Polypeptides  
 ; FILE REFERENCE: 002.00220 (MEDY/P24958PC)  
 ; CURRENT APPLICATION NUMBER: US/10/343,663A  
 ; CURRENT FILING DATE: 2003-12-03  
 ; PRIOR APPLICATION NUMBER: PCT/GB01/03481  
 ; PRIOR FILING DATE: 2001-08-03  
 ; PRIOR APPLICATION NUMBER: GB 0018908.4  
 ; PRIOR FILING DATE: 2000-08-03  
 ; NUMBER OF SEQ ID NOS: 68  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 36  
 ; LENGTH: 21  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Peptide  
 US-10-343-663A-36

Query Match 0.7%; Score 7; DB 16; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 YTSRAGE 200  
 |||||  
 DB 2 YTSRAGE 8

RESULT 97  
 US-09-864-761-45902

; Sequence 45902, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecm1ca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 45902  
; LENGTH: 41  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC008521.5  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.71  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.77  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.66  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.73  
US-09-864-761-45902

Query Match 0.7%; Score 7; DB 9; Length 41;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 945 EKSSSSE 951  
Db 32 EKSSSSE 38  
|||||

RESULT 98

US-10-424-599-179769  
; Sequence 179769, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 179769  
; LENGTH: 46  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_133347C.1.pep  
US-10-424-599-179769

Query Match 0.7%; Score 7; DB 15; Length 46;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 779 VLAYNNS 785  
Db 4 VLAYNNS 10  
|||||

RESULT 99

US-10-424-599-185551  
; Sequence 185551, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 185551  
; LENGTH: 47  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_138567C.1.pep  
US-10-424-599-185551

Query Match 0.7%; Score 7; DB 15; Length 47;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 718 GKALLSK 724  
Db 22 GKALLSK 28  
|||||

RESULT 100

US-10-424-599-225766  
; Sequence 225766, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ.ID NOS: 285684
; SEQ ID NO 225766
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_45899C.1.pap
US-10-424-599-225766

```

```

Query Match      0.7%; Score 7; DB 15; Length 51;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      427 LQKVKRN 433
      |||||
Db      19 LQKVKRN 25

```

Search completed: August 28, 2005, 11:22:10  
Job time : 174 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2005, 11:02:48 ; Search time 43 Seconds  
(without alignments)  
1749.913 Million cell updates/sec

Title: US-10-078-531-2

Perfect score: 1008

Sequence: 1 MKKHLKTLVALTLTVSVVTH.....LGYTSVALLSLTAIKKKY 1008

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 513545 seqs, 74649064 residues

Word size : 0

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 110 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	0.9	557	4	US-09-568-816A-2
2	9	0.9	907	4	US-09-170-496D-264
3	9	0.9	907	4	US-09-170-496D-278
4	8	0.8	350	2	US-08-852-481-2
5	8	0.8	364	4	US-09-328-352-4367
6	8	0.8	373	4	US-09-248-796A-19382
7	8	0.8	531	1	US-08-531-601-1
8	8	0.8	531	2	US-08-859-032-1
9	8	0.8	535	1	US-08-737-597-10
10	8	0.8	555	4	US-09-227-421-2
11	8	0.8	555	4	US-09-479-855-2
12	8	0.8	688	2	US-09-016-000-2
13	8	0.8	976	3	US-08-560-005-2
14	8	0.8	976	3	US-09-195-868-14
15	8	0.8	976	3	US-09-418-540-2
16	8	0.8	976	4	US-09-969-528-2
17	8	0.8	1187	3	US-08-664-962B-8
18	8	0.8	1187	3	US-09-311-743-8
19	8	0.8	1189	3	US-09-195-868-15
20	8	0.8	1229	3	US-09-195-868-28
21	8	0.8	3878	4	US-09-914-259-11
22	7	0.7	16	3	US-09-057-052-8
23	7	0.7	16	3	US-09-057-052-12
24	7	0.7	16	4	US-09-695-466-8
25	7	0.7	16	4	US-09-695-466-12
26	7	0.7	26	4	US-09-270-767-60016
27	7	0.7	38	4	US-09-270-767-44565

28	7	0.7	39	4	US-09-270-767-61674	Sequence 61674, A
29	7	0.7	60	4	US-09-621-976-6627	Sequence 6627, Ap
30	7	0.7	60	4	US-09-621-976-6631	Sequence 6631, Ap
31	7	0.7	62	4	US-09-248-796A-20152	Sequence 20152, A
32	7	0.7	77	4	US-09-513-999C-5429	Sequence 5429, Ap
33	7	0.7	81	4	US-09-621-976-6626	Sequence 6626, Ap
34	7	0.7	81	4	US-09-621-976-6628	Sequence 6628, Ap
35	7	0.7	81	4	US-09-621-976-6629	Sequence 6629, Ap
36	7	0.7	81	4	US-09-621-976-6630	Sequence 6630, Ap
37	7	0.7	85	4	US-09-248-796A-19130	Sequence 19130, A
38	7	0.7	87	4	US-09-248-796A-16501	Sequence 16501, A
39	7	0.7	90	4	US-09-270-767-40137	Sequence 40137, A
40	7	0.7	90	4	US-09-270-767-55353	Sequence 55353, A
41	7	0.7	94	4	US-09-270-767-44382	Sequence 44382, A
42	7	0.7	101	4	US-09-902-540-11181	Sequence 11181, A
43	7	0.7	105	4	US-09-513-999C-5997	Sequence 5997, Ap
44	7	0.7	106	4	US-09-270-767-44049	Sequence 44049, A
45	7	0.7	111	4	US-09-248-796A-20841	Sequence 20841, A
46	7	0.7	120	4	US-09-071-035-206	Sequence 206, App
47	7	0.7	120	4	US-09-071-035-210	Sequence 210, App
48	7	0.7	121	4	US-09-252-991A-28777	Sequence 28777, A
49	7	0.7	123	4	US-09-270-767-33193	Sequence 33193, A
50	7	0.7	123	4	US-09-270-767-48410	Sequence 48410, A
51	7	0.7	136	4	US-09-270-767-57861	Sequence 57861, A
52	7	0.7	143	4	US-09-134-000C-6044	Sequence 6044, Ap
53	7	0.7	144	3	US-08-858-207A-523	Sequence 523, App
54	7	0.7	147	4	US-09-134-000C-6221	Sequence 6221, Ap
55	7	0.7	152	3	US-08-679-493A-203	Sequence 203, App
56	7	0.7	161	4	US-09-252-991A-24022	Sequence 24022, A
57	7	0.7	172	4	US-09-538-092-450	Sequence 450, App
58	7	0.7	176	4	US-09-270-767-38641	Sequence 38641, A
59	7	0.7	176	4	US-09-270-767-53858	Sequence 53858, A
60	7	0.7	176	4	US-09-893-737-22	Sequence 22, Appl
61	7	0.7	180	4	US-09-270-767-35060	Sequence 35060, A
62	7	0.7	180	4	US-09-270-767-50277	Sequence 50277, A
63	7	0.7	181	4	US-09-270-767-61846	Sequence 61846, A
64	7	0.7	190	4	US-09-902-540-15188	Sequence 15188, A
65	7	0.7	208	4	US-09-270-767-46119	Sequence 46119, A
66	7	0.7	211	4	US-09-248-796A-15823	Sequence 15823, A
67	7	0.7	215	3	US-09-305-984-18	Sequence 18, Appl
68	7	0.7	215	3	US-09-305-984-64	Sequence 64, Appl
69	7	0.7	215	4	US-09-073-541A-18	Sequence 18, Appl
70	7	0.7	215	4	US-09-493-940-18	Sequence 18, Appl
71	7	0.7	215	4	US-09-493-940-64	Sequence 64, Appl
72	7	0.7	215	4	US-09-583-110-5320	Sequence 5320, Ap
73	7	0.7	219	4	US-09-107-433-4810	Sequence 4810, Ap
74	7	0.7	224	3	US-09-305-984-72	Sequence 72, Appl
75	7	0.7	224	3	US-09-305-984-74	Sequence 74, Appl
76	7	0.7	224	3	US-09-305-984-76	Sequence 76, Appl
77	7	0.7	224	4	US-09-493-940-72	Sequence 72, Appl
78	7	0.7	224	4	US-09-493-940-74	Sequence 74, Appl
79	7	0.7	224	4	US-09-493-940-76	Sequence 76, Appl
80	7	0.7	236	4	US-09-902-540-11067	Sequence 11067, A
81	7	0.7	239	4	US-09-198-452A-399	Sequence 399, App
82	7	0.7	239	4	US-09-438-185A-384	Sequence 384, App
83	7	0.7	241	2	US-08-867-087B-11	Sequence 11, Appl
84	7	0.7	241	4	US-09-107-532A-5876	Sequence 5876, Ap
85	7	0.7	242	4	US-09-252-991A-21887	Sequence 21887, A
86	7	0.7	243	4	US-09-328-352-4519	Sequence 4519, Ap
87	7	0.7	248	4	US-09-252-991A-32999	Sequence 32999, A
88	7	0.7	250	4	US-09-270-767-40859	Sequence 40859, A
89	7	0.7	250	4	US-09-270-767-56075	Sequence 56075, A
90	7	0.7	256	4	US-09-248-796A-17359	Sequence 17359, A
91	7	0.7	259	4	US-09-489-039A-7271	Sequence 7271, Ap
92	7	0.7	261	4	US-09-489-039A-11968	Sequence 11968, A
93	7	0.7	268	4	US-09-270-767-42555	Sequence 42555, A
94	7	0.7	270	4	US-09-248-796A-19421	Sequence 19421, A
95	7	0.7	272	4	US-09-489-039A-10410	Sequence 10410, A
96	7	0.7	289	2	US-08-741-437-1	Sequence 1, Appli
97	7	0.7	289	2	US-09-134-593-1	Sequence 1, Appli
98	7	0.7	289	4	US-09-248-796A-18420	Sequence 18420, A
99	7	0.7	289	4	US-09-538-092-1335	Sequence 1335, Ap
100	7	0.7	292	4	US-09-535-521-2	Sequence 2, Appli

101 7 0.7 292 4 US-09-535-521-5 Sequence 5, Appli  
102 7 0.7 294 4 US-09-252-991A-26953 Sequence 26953, A  
103 7 0.7 295 4 US-09-328-352-4378 Sequence 4378, Ap  
104 7 0.7 302 4 US-09-252-991A-21875 Sequence 21875, A  
105 7 0.7 306 4 US-09-489-039A-10023 Sequence 10023, A  
106 7 0.7 307 4 US-09-489-039A-10116 Sequence 10116, A  
107 7 0.7 313 3 US-08-985-335-8 Sequence 8, Appli  
108 7 0.7 313 3 US-09-410-372-8 Sequence 8, Appli  
109 7 0.7 314 2 US-08-822-701-10 Sequence 10, Appli  
110 7 0.7 314 3 US-08-935-855-10 Sequence 10, Appli

## ALIGNMENTS

RESULT 1  
US-09-568-816A-2  
; Sequence 2, Application US/09568816A  
; Patent No. 6440599  
; GENERAL INFORMATION:  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Swedlund, Brad  
; APPLICANT: Simard, Jacques  
; APPLICANT: Rommens, Johanna M.  
; APPLICANT: Myriad Genetics, Inc.  
; APPLICANT: Hospital for Sick Children  
; TITLE OF INVENTION: CA7 CG04 Gene  
; FILE REFERENCE: 2318-237-II  
; CURRENT APPLICATION NUMBER: US/09/568,816A  
; PRIOR FILING DATE: 2000-05-11  
; PRIOR APPLICATION NUMBER: US 60/134,209  
; PRIOR FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 557  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-568-816A-2

Query Match 0.9%; Score 9; DB 4; Length 557;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSSSS 952  
| | | | | | | |  
DB 18 SEKSSSSS 26

RESULT 2  
US-09-170-496D-264  
; Sequence 264, Application US/09170496D  
; Patent No. 6555339  
; GENERAL INFORMATION:  
; APPLICANT: Behan, Dominic P.  
; APPLICANT: Chalmers, Derek T.  
; APPLICANT: Liaw, Chen W.  
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-  
; FILE REFERENCE: AREN-0040  
; CURRENT APPLICATION NUMBER: US/09/170,496D  
; CURRENT FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 294  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 264  
; LENGTH: 907  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-170-496D-264

Query Match 0.9%; Score 9; DB 4; Length 907;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVOAP 600  
| | | | | | | |  
DB 174 LTEIPVOAP 182

## RESULT 3

US-09-170-496D-278  
; Sequence 278, Application US/09170496D  
; Patent No. 6555339  
; GENERAL INFORMATION:  
; APPLICANT: Behan, Dominic P.  
; APPLICANT: Chalmers, Derek T.  
; APPLICANT: Liaw, Chen W.  
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-  
; FILE REFERENCE: AREN-0040  
; CURRENT APPLICATION NUMBER: US/09/170,496D  
; CURRENT FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 294  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 278  
; LENGTH: 907  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-170-496D-278

Query Match 0.9%; Score 9; DB 4; Length 907;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 LTEIPVOAP 600  
| | | | | | | |  
DB 174 LTEIPVOAP 182

## RESULT 4

US-08-852-481-2  
; Sequence 2, Application US/08852481  
; Patent No. 5928931  
; GENERAL INFORMATION:  
; APPLICANT: Grun Ph.D., Felix  
; APPLICANT: Buck Ph.D., Jochen  
; APPLICANT: Hammerling Ph.D., Ulrich  
; TITLE OF INVENTION: ISOLATION, PURIFICATION AND CLONING OF  
; TITLE OF INVENTION: RETINOL DEHYDRATASE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,481  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/017,178  
; FILING DATE: 09-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/771  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600



; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 350 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-852-481-2

Query Match 0.8%; Score 8; DB 2; Length 350;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 668 LADGEHFI 675  
Db 303 LADGEHFI 310  
|||||

RESULT 5  
US-09-328-352-4367  
; Sequence 4367, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4367  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4367

Query Match 0.8%; Score 8; DB 4; Length 364;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 997 LSLITAIK 1004  
Db 164 LSLITAIK 171  
|||||

RESULT 6  
US-09-248-796A-19382  
; Sequence 19382, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 19382  
; LENGTH: 373  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-19382

Query Match 0.8%; Score 8; DB 4; Length 373;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 LEKELDLL 802  
|||||

Db 38 LEKELDLL 45

RESULT 7  
US-08-531-601-1  
; Sequence 1, Application US/08531601  
; Patent No. 568684  
; GENERAL INFORMATION:  
; APPLICANT: YOSHIGI, NAOHIRO  
; APPLICANT: MAEBA, HIDEO  
; APPLICANT: OKADA, YUKIO  
; TITLE OF INVENTION: RECOMBINANT BETA-AMYLASE  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,601  
; FILING DATE: 21-SEP-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP HEI 6-233086  
; FILING DATE: 28-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F  
; REGISTRATION NUMBER: 08/531,601  
; REFERENCE/DOCKET NUMBER: 2589-031-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 531 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-531-601-1

Query Match 0.8%; Score 8; DB 1; Length 531;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVECKGP 811  
Db 51 GLVECKGP 58  
|||||

RESULT 8  
US-08-859-032-1  
; Sequence 1, Application US/08859032  
; Patent No. 5863784  
; GENERAL INFORMATION:  
; APPLICANT: YOSHIGI, NAOHIRO  
; APPLICANT: MAEBA, HIDEO  
; APPLICANT: OKADA, YUKIO  
; TITLE OF INVENTION: RECOMBINANT BETA-AMYLASE  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VIRGINIA

;; COUNTRY: USA  
;; ZIP: 22202  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/859,032  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/531,601  
;; FILING DATE: 21-SEP-1995  
;; APPLICATION NUMBER: JP HEI 6-233086  
;; FILING DATE: 28-SEP-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: OBLON, NORMAN F.  
;; REGISTRATION NUMBER: 08/531,601  
;; REFERENCE/DOCKET NUMBER: 2589-031-0  
;; TELEPHONE: (703) 413-3000  
;; TELEFAX: (703) 413-2220  
;; TELEX: 248855 OPAT UR  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 531 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-859-032-1

Query Match 0.8%; Score 8; DB 2; Length 531;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811  
|||  
DB 51 GLVEGKGP 58

RESULT 9  
US-08-737-597-10  
; Sequence 10, Application US/08737597  
; Patent No. 572657  
; GENERAL INFORMATION:  
; APPLICANT: YOSHIGI, NAOHICO  
; TITLE OF INVENTION: BARLEY BETA AMYLASE STRUCTURAL GENE  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/737,597  
; FILING DATE: 25-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP96/00799  
; FILING DATE: 27-MAR-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7/92004

;; FILING DATE: 25-MAR-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: OBLON, NORMAN F.  
;; REGISTRATION NUMBER: 24,618  
;; REFERENCE/DOCKET NUMBER: 2589-044-0 PCT  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 703-413-3000  
;; TELEFAX: 703-413-2220  
;; INFORMATION FOR SEQ ID NO: 10:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 535 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-737-597-10

Query Match 0.8%; Score 8; DB 1; Length 535;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 GLVEGKGP 811  
|||  
DB 55 GLVEGKGP 62

RESULT 10  
US-09-227-421-2  
; Sequence 2, Application US/09227421  
; Patent No. 6559357  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Mizukami, Yukiko  
; TITLE OF INVENTION: Methods for Altering Organ Mass, Controlling Fertility  
; TITLE OF INVENTION: and Enhancing Asexual Reproduction in Plants  
; FILE REFERENCE: 023070-090700PC  
; CURRENT APPLICATION NUMBER: US/09/227,421  
; CURRENT FILING DATE: 1999-01-08  
; PRIOR APPLICATION NUMBER: US 09/227,421  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 555  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-227-421-2

Query Match 0.8%; Score 8; DB 4; Length 555;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVALTIL 12  
|||  
DB 536 LKTVALTIL 543

RESULT 11  
US-09-479-855-2  
; Sequence 2, Application US/09479855  
; Patent No. 6639128  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Mizukami, Yukiko  
; TITLE OF INVENTION: Methods for Altering Organ Mass, Controlling Fertility  
; TITLE OF INVENTION: and Enhancing Asexual Reproduction in Plants  
; FILE REFERENCE: 023070-0907200S  
; CURRENT APPLICATION NUMBER: US/09/479,855  
; CURRENT FILING DATE: 2000-01-07  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2  
; LENGTH: 555  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; OTHER INFORMATION: AINTEGUMENTA (ANT)  
US-09-479-855-2

Query Match 0.8%; Score 8; DB 4; Length 555;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LKTVLTL 12  
DB 536 LKTVLTL 543

RESULT 12  
US-09-016-000-2  
; Sequence 2, Application US/09016000  
; Patent No. 5962232  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Bandman, Olga  
; APPLICANT: Akerblom, Ingrid E.  
; APPLICANT: Shah, Purvi  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl G.  
; TITLE OF INVENTION: PROTEIN KINASE MOLECULES  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,000  
; FILING DATE: HERewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0465 US  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 688 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: HEARNOT01  
; CLONE: 307624  
US-09-016-000-2

Query Match 0.8%; Score 8; DB 2; Length 688;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 944 SEKSSES 951

DB 5 SEKSSES 12

RESULT 13  
US-08-560-005-2  
; Sequence 2, Application US/08560005  
; Patent No. 6001354  
; GENERAL INFORMATION:  
; APPLICANT: Pot, David A.  
; APPLICANT: Williams, Lewis T.  
; APPLICANT: Jefferson, Anne Bennett  
; APPLICANT: Majerus, Philip W.  
; TITLE OF INVENTION: No. 6001354el Grb2 Associating Protein and Nucleic  
; TITLE OF INVENTION: Acids Encoding Therefor  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew  
; STREET: One Market Plaza, Steuart Tower, Suite 2000  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/560,005  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dow, Karen B.  
; REGISTRATION NUMBER: 29,684  
; REFERENCE/DOCKET NUMBER: 2307K-0624000  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 976 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-560-005-2

Query Match 0.8%; Score 8; DB 3; Length 976;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 SKLSQLS 908  
DB 50 SKLSQLS 57

RESULT 14  
US-09-195-868-14  
; Sequence 14, Application US/09195868  
; Patent No. 6090621  
; GENERAL INFORMATION:  
; APPLICANT: KAVANAUGH MD, MICHAEL  
; APPLICANT: POT PH.D., DAVID  
; APPLICANT: WILLIAMS MDPHD, LEWIS T.  
; TITLE OF INVENTION: SIGNALING INOSITOL POLYPHOSPHATE  
; TITLE OF INVENTION: 5-PHOSPHATASES (SIPs)  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CHIRON CORPORATION  
; STREET: 4560 HORTON STREET  
; CITY: EMERYVILLE  
; STATE: CA  
; COUNTRY: USA

REGISTRATION NUMBER: 29,684  
REFERENCE/DOCKET NUMBER: 2307K-0624000  
TELECOMMUNICATION INFORMATION:

RESULT 17  
US-08-664-962B-8  
; Sequence 8, Application US/08664962B  
; Patent No. 6218162  
; GENERAL INFORMATION:  
; APPLICANT: Krystal, Gerald  
; TITLE OF INVENTION: SH2-CONTAINING INOSITOL-PHOSPHATASE  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MERCHANT & GOULD  
; STREET: 3100 No. 6218162 West Center, 90 South Seventh Street  
; CITY: Minneapolis  
; STATE: Minnesota  
; COUNTRY: U.S.A.  
; ZIP: 55402-4131  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/664,962B  
; FILING DATE: 14-JUN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mueller, Douglas P.  
; REGISTRATION NUMBER: 30,300  
; REFERENCE/DOCKET NUMBER: M&G 7933.49-US-01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-332-5300  
; TELEFAX: 612-332-9081  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1187 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-664-962B-8

Query Match 0.8%; Score 8; DB 3; Length 1187;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 901 SKLSOLTS 908  
Db 261 SKLSOLTS 268

RESULT 18  
US-09-311-743-8  
; Sequence 8, Application US/09311743  
; Patent No. 6238903  
; GENERAL INFORMATION:  
; APPLICANT: Krystal, Gerald  
; TITLE OF INVENTION: SH2-CONTAINING INOSITOL-PHOSPHATASE  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BERESKIN & PARR  
; STREET: 40 King Street West  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5H 3Y2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/311,743  
; FILING DATE: 14-May-1999

CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gravelle, Micheline  
; REGISTRATION NUMBER: 40,261  
; REFERENCE/DOCKET NUMBER: 7771-32  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 416-364-7311  
; TELEFAX: 416-361-1398  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1187 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-311-743-8

Query Match 0.8%; Score 8; DB 3; Length 1187;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 901 SKLSOLTS 908  
Db 261 SKLSOLTS 268

RESULT 19  
US-09-195-868-15  
; Sequence 15, Application US/09195868  
; Patent No. 6090621  
; GENERAL INFORMATION:  
; APPLICANT: KAVANAUGH MD, MICHAEL  
; APPLICANT: POT PH.D., DAVID  
; APPLICANT: WILLIAMS MDPHD, LEWIS T.  
; TITLE OF INVENTION: SIGNALING INOSITOL POLYPHOSPHATE  
; TITLE OF INVENTION: 5-PHOSPHATASES (SIPs)  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CHIRON CORPORATION  
; STREET: 4560 HORTON STREET  
; CITY: EMERYVILLE  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/195,868  
; FILING DATE:  
; CLASSIFICATION:  
; APPLICATION NUMBER: US/09/195,868  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FIRESTONE, LEIGH H.  
; REGISTRATION NUMBER: 36,831  
; REFERENCE/DOCKET NUMBER: 1182.004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-923-2707  
; TELEFAX: 510-655-3542  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1189 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-195-868-15

Query Match 0.8%; Score 8; DB 3; Length 1189;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 SKLSQTS 908  
Db 263 SKLSQTS 270

## RESULT 20

US-09-195-868-28  
; Sequence 28, Application US/09195868  
; Patent No. 6090621  
; GENERAL INFORMATION:  
; APPLICANT: KAVANAUGH MD, MICHAEL  
; APPLICANT: POT PH.D., DAVID  
; APPLICANT: WILLIAMS MDPHD, LEWIS T.  
; TITLE OF INVENTION: SIGNALING INOSITOL POLYPHOSPHATE  
; TITLE OF INVENTION: 5-PHOSPHATASES (SIPs)  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CHIRON CORPORATION  
; STREET: 4560 HORTON STREET  
; CITY: EMERYVILLE  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/195,868  
; FILING DATE:  
; CLASSIFICATION:  
; APPLICATION NUMBER: US/09/195,868  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FIRESTONE, LEIGH H.  
; REGISTRATION NUMBER: 36,831  
; REFERENCE/DOCKET NUMBER: 1182.004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-923-2707  
; TELEFAX: 510-655-3542  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1229 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-195-868-28

Query Match 0.8%; Score 8; DB 3; Length 1229;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 SKLSQTS 908  
Db 303 SKLSQTS 310

## RESULT 21

US-09-914-259-11  
; Sequence 11, Application US/09914259  
; Patent No. 6495336  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; APPLICANT: Williams, Mark  
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
; FILE REFERENCE: 8471-010-999  
; CURRENT APPLICATION NUMBER: US/09/914,259  
; CURRENT FILING DATE: 2000-11-21

; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 3878  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-914-259-11

Query Match 0.8%; Score 8; DB 4; Length 3878;  
Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLEKDKL 523  
Db 357 TLEKDKL 364

## RESULT 22

US-09-057-052-8  
; Sequence 8, Application US/09057052  
; Patent No. 6331422  
; GENERAL INFORMATION:  
; APPLICANT: Hubbell, Jeffrey A.  
; APPLICANT: Schense, Jason  
; TITLE OF INVENTION: Enzyme-Mediated Modification of Fibrin for  
; TITLE OF INVENTION: Tissue Engineering  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Vinson & Elkins L.L.P.  
; STREET: 600 Congress Avenue, Suite 2700  
; CITY: Austin  
; STATE: Texas  
; COUNTRY: US  
; ZIP: 78701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/057,052  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/042,143  
; FILING DATE: 03-APR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mayfield, Denise L.  
; REGISTRATION NUMBER: 33,732  
; REFERENCE/DOCKET NUMBER: CAL430.23000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 495-8400  
; TELEFAX: (512) 495-8612  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 16 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-09-057-052-8

Query Match 0.7%; Score 7; DB 3; Length 16;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 855 DTIGEGQ 861  
Db 4 DTIGEGQ 10

## RESULT 23

US-09-057-052-12  
; Sequence 12, Application US/09057052

Patent No. 6331422  
GENERAL INFORMATION:  
APPLICANT: Hubbell, Jeffrey A.  
APPLICANT: Schense, Jason  
TITLE OF INVENTION: Enzyme-Mediated Modification of Fibrin for  
Tissue Engineering  
TITLE OF INVENTION: Tissue Engineering  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Vinson & Elkins L.L.P.  
STREET: 600 Congress Avenue, Suite 2700  
CITY: Austin  
STATE: Texas  
COUNTRY: US  
ZIP: 78701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/057,052  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/042,143  
FILING DATE: 03-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mayfield, Denise L.  
REGISTRATION NUMBER: 33,732  
REFERENCE/DOCKET NUMBER: CAL430.23000  
TELEPHONE: (512) 495-8400  
TELEFAX: (512) 495-8612  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY: linear  
US-09-057-052-12

Query Match 0.7%; Score 7; DB 3; Length 16;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 855 DTIGEGQ 861  
Db 4 DTIGEGQ 10

RESULT 24  
US-09-695-466-8  
Sequence 8, Application US/09695466  
Patent No. 6607740  
GENERAL INFORMATION:  
APPLICANT: Hubbell, Jeffrey A.  
APPLICANT: Schense, Jason  
TITLE OF INVENTION: Enzyme-Mediated Modification of Fibrin for  
Tissue Engineering  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Vinson & Elkins L.L.P.  
STREET: 600 Congress Avenue, Suite 2700  
CITY: Austin  
STATE: Texas  
COUNTRY: US  
ZIP: 78701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/695,466  
FILING DATE: 24-Oct-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/057,052  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 60/042,143  
FILING DATE: 03-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mayfield, Denise L.  
REGISTRATION NUMBER: 33,732  
REFERENCE/DOCKET NUMBER: CAL430.23000  
TELEPHONE: (512) 495-8400  
TELEFAX: (512) 495-8612  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-695-466-8

Query Match 0.7%; Score 7; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 855 DTIGEGQ 861  
Db 4 DTIGEGQ 10

RESULT 25  
US-09-695-466-12  
Sequence 12, Application US/09695466  
Patent No. 6607740  
GENERAL INFORMATION:  
APPLICANT: Hubbell, Jeffrey A.  
APPLICANT: Schense, Jason  
TITLE OF INVENTION: Enzyme-Mediated Modification of Fibrin for  
Tissue Engineering  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Vinson & Elkins L.L.P.  
STREET: 600 Congress Avenue, Suite 2700  
CITY: Austin  
STATE: Texas  
COUNTRY: US  
ZIP: 78701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/695,466  
FILING DATE: 24-Oct-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/057,052  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 60/042,143  
FILING DATE: 03-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mayfield, Denise L.  
REGISTRATION NUMBER: 33,732  
REFERENCE/DOCKET NUMBER: CAL430.23000  
TELEPHONE: (512) 495-8400  
TELEFAX: (512) 495-8612  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:



```
;
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-695-466-12

Query Match          0.7%; Score 7; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 855 DTIGEGQ 861
Db 4 DTIGEGQ 10

RESULT 26
US-09-270-767-60016
; Sequence 60016, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60016
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-60016

Query Match          0.7%; Score 7; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 825 KTPLPLP 831
Db 12 KTPLPLP 18

RESULT 27
US-09-270-767-44565
; Sequence 44565, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44565
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-44565

Query Match          0.7%; Score 7; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 825 KTPLPLP 831
Db 24 KTPLPLP 30

RESULT 28
US-09-270-767-61674
; Sequence 61674, Application US/09270767
```

```
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61674
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-61674

Query Match          0.7%; Score 7; DB 4; Length 39;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 ASSSISG 44
Db 11 ASSSISG 17

RESULT 29
US-09-621-976-6627
; Sequence 6627, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.Pm
; SEQ ID NO 6627
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6627

Query Match          0.7%; Score 7; DB 4; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 808 GKGPLAQ 814
Db 28 GKGPLAQ 34

RESULT 30
US-09-621-976-6631
; Sequence 6631, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.Pm
; SEQ ID NO 6631
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6631
```

```
Query Match      0.7%; Score 7; DB 4; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      808 GKGPLAQ 814
DB      28 GKGPLAQ 34

RESULT 31
US-09-248-796A-20152
; Sequence 20152, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20152
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20152

Query Match      0.7%; Score 7; DB 4; Length 62;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      994 VALLSLI 1000
DB      12 VALLSLI 18

RESULT 32
US-09-513-999C-5429
; Sequence 5429, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1998-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 5429
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: UNSURE
; LOCATION: 73
; OTHER INFORMATION: Xaa=Glu or Gly or Lys or Arg
US-09-513-999C-5429

Query Match      0.7%; Score 7; DB 4; Length 77;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      995 ALLSLIT 1001
DB      11 ALLSLIT 1001

Query Match      0.7%; Score 7; DB 4; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      808 GKGPLAQ 814
DB      28 GKGPLAQ 34

RESULT 33
US-09-621-976-6626
; Sequence 6626, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 6626
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6626

Query Match      0.7%; Score 7; DB 4; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      808 GKGPLAQ 814
DB      49 GKGPLAQ 55

RESULT 34
US-09-621-976-6628
; Sequence 6628, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 6628
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6628

Query Match      0.7%; Score 7; DB 4; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      808 GKGPLAQ 814
DB      49 GKGPLAQ 55

RESULT 35
US-09-621-976-6629
; Sequence 6629, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
```

! NUMBER OF SEQ ID NOS: 19335  
! SOFTWARE: Patent.pm  
! SEQ ID NO 6629  
! LENGTH: 81  
! TYPE: PRT  
! ORGANISM: Homo sapiens  
US-09-621-976-6629

Query Match 0.7%; Score 7; DB 4; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 808 KGKPLAQ 814  
|||||  
Db 49 KGKPLAQ 55

## RESULT 36

US-09-621-976-6630  
! Sequence 6630, Application US/09621976  
! Patent No. 6639063  
! GENERAL INFORMATION:  
! APPLICANT: Dumas Milne Edwards, J.B.  
! APPLICANT: Jobert, S.  
! APPLICANT: Giordano, J.Y.  
! TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
! FILE REFERENCE: GENSET.054PR2  
! CURRENT APPLICATION NUMBER: US/09/621,976  
! CURRENT FILING DATE: 2000-07-21  
! NUMBER OF SEQ ID NOS: 19335  
! SOFTWARE: Patent.pm  
! SEQ ID NO 6630  
! LENGTH: 81  
! TYPE: PRT  
! ORGANISM: Homo sapiens  
! FEATURE:  
! NAME/KEY: UNSURE  
! LOCATION: 36  
! OTHER INFORMATION: Xaa = Asp, Glu, Val  
US-09-621-976-6630

Query Match 0.7%; Score 7; DB 4; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 808 KGKPLAQ 814  
|||||  
Db 49 KGKPLAQ 55

## RESULT 37

US-09-248-796A-19130  
! Sequence 19130, Application US/09248796A  
! Patent No. 6747137  
! GENERAL INFORMATION:  
! APPLICANT: Keith Weinstock et al

! TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
! FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS  
! FILE REFERENCE: 107196.132  
! CURRENT APPLICATION NUMBER: US/09/248,796A  
! CURRENT FILING DATE: 1999-02-12  
! PRIOR APPLICATION NUMBER: US 60/074,725  
! PRIOR FILING DATE: 1998-02-13  
! PRIOR APPLICATION NUMBER: US 60/096,409  
! PRIOR FILING DATE: 1998-08-13  
! NUMBER OF SEQ ID NOS: 28208  
! SEQ ID NO 19130  
! LENGTH: 85  
! TYPE: PRT  
! ORGANISM: Candida albicans  
US-09-248-796A-19130

Query Match 0.7%; Score 7; DB 4; Length 85;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 769 QLLERSI 775  
|||||  
Db 77 QLLERSI 83

## RESULT 38

US-09-248-796A-16501  
! Sequence 16501, Application US/09248796A  
! Patent No. 6747137  
! GENERAL INFORMATION:  
! APPLICANT: Keith Weinstock et al  
! TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
! FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS  
! FILE REFERENCE: 107196.132  
! CURRENT APPLICATION NUMBER: US/09/248,796A  
! CURRENT FILING DATE: 1999-02-12  
! PRIOR APPLICATION NUMBER: US 60/074,725  
! PRIOR FILING DATE: 1998-02-13  
! PRIOR APPLICATION NUMBER: US 60/096,409  
! PRIOR FILING DATE: 1998-08-13  
! NUMBER OF SEQ ID NOS: 28208  
! SEQ ID NO 16501  
! LENGTH: 87  
! TYPE: PRT  
! ORGANISM: Candida albicans  
US-09-248-796A-16501

Query Match 0.7%; Score 7; DB 4; Length 87;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 692 LKLI EGL 698  
|||||  
Db 12 LKLI EGL 18

## RESULT 39

US-09-270-767-40137  
! Sequence 40137, Application US/09270767  
! Patent No. 6703491  
! GENERAL INFORMATION:  
! APPLICANT: Homburger et al.  
! TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
! FILE REFERENCE: File Reference: 7326-094  
! CURRENT APPLICATION NUMBER: US/09/270,767  
! CURRENT FILING DATE: 1999-03-17  
! NUMBER OF SEQ ID NOS: 62517  
! SOFTWARE: PatentIn Ver. 2.0  
! SEQ ID NO 40137  
! LENGTH: 90  
! TYPE: PRT  
! ORGANISM: Drosophila melanogaster  
US-09-270-767-40137

Query Match 0.7%; Score 7; DB 4; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 225 LLKKVTI 231  
|||||  
Db 25 LLKKVTI 31

## RESULT 40

US-09-270-767-55353  
! Sequence 55353, Application US/09270767  
! Patent No. 6703491  
! GENERAL INFORMATION:  
! APPLICANT: Homburger et al.  
! TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

FILE REFERENCE: File Reference: 7326-094  
CURRENT APPLICATION NUMBER: US/09/270,767  
CURRENT FILING DATE: 1999-03-17  
NUMBER OF SEQ ID NOS: 62517  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 55353  
LENGTH: 90  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-09-270-767-55353

Query Match 0.7%; Score 7; DB 4; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 LKKVTI 231  
|||||  
Db 25 LKKVTI 31

RESULT 41  
US-09-270-767-44382  
Sequence 44382, Application US/09270767  
Patent No. 6703491  
GENERAL INFORMATION:  
APPLICANT: Homburger et al.  
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

FILE REFERENCE: File Reference: 7326-094  
CURRENT APPLICATION NUMBER: US/09/270,767  
CURRENT FILING DATE: 1999-03-17  
NUMBER OF SEQ ID NOS: 62517  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 44382  
LENGTH: 94  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-09-270-767-44382

Query Match 0.7%; Score 7; DB 4; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 KDLAKG 90  
|||||  
Db 26 KDLAKG 32

RESULT 42  
US-09-902-540-11181  
Sequence 11181, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 11181  
LENGTH: 101  
TYPE: PRT  
ORGANISM: Myxococcus xanthus  
US-09-902-540-11181

Query Match 0.7%; Score 7; DB 4; Length 101;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 ALVTCKA 764  
|||||  
Db 69 ALVTCKA 75  
RESULT 43  
US-09-513-999C-5997  
Sequence 5997, Application US/09513999C  
Patent No. 6783961  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Duclert, A.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
Patent No. 6783961  
FILE REFERENCE: 59 US2, REG  
CURRENT APPLICATION NUMBER: US/09/513,999C  
CURRENT FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/122,487  
PRIOR FILING DATE: 1999-02-26  
NUMBER OF SEQ ID NOS: 36681  
SOFTWARE: Patent.pm  
SEQ ID NO 5997  
LENGTH: 105  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-513-999C-5997

Query Match 0.7%; Score 7; DB 4; Length 105;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 LVGLSKS 152  
|||||  
Db 10 LVGLSKS 16

RESULT 44  
US-09-270-767-44049  
Sequence 44049, Application US/09270767  
Patent No. 6703491  
GENERAL INFORMATION:  
APPLICANT: Homburger et al.  
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

FILE REFERENCE: File Reference: 7326-094  
CURRENT APPLICATION NUMBER: US/09/270,767  
CURRENT FILING DATE: 1999-03-17  
NUMBER OF SEQ ID NOS: 62517  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 44049  
LENGTH: 106  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-09-270-767-44049

Query Match 0.7%; Score 7; DB 4; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 946 KSSSES 952  
|||||  
Db 90 KSSSES 96

RESULT 45  
US-09-248-796A-20841  
Sequence 20841, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

FILE REFERENCE: 107196.132  
FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 20841  
; LENGTH: 111  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-20841

Query Match 0.7%; Score 7; DB 4; Length 111;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 996 LLSLITA 1002  
|||||||  
Db 32 LLSLITA 38

RESULT 46  
US-09-071-035-206  
; Sequence 206, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 206:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-071-035-206

Query Match 0.7%; Score 7; DB 4; Length 120;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 975 LPRTGSK 981  
|||||||  
Db 85 LPRTGSK 91

US-09-071-035-206  
; Sequence 206, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 206:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-071-035-206

Query Match 0.7%; Score 7; DB 4; Length 120;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 975 LPRTGSK 981  
|||||||  
Db 85 LPRTGSK 91

RESULT 47  
US-09-071-035-210  
; Sequence 210, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 210:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-071-035-210

Query Match 0.7%; Score 7; DB 4; Length 120;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 975 LPRTGSK 981  
|||||||  
Db 85 LPRTGSK 91

RESULT 48  
US-09-252-991A-28777  
; Sequence 28777, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28777  
; LENGTH: 121  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa

## US-09-252-991A-28777

Query Match 0.7%; Score 7; DB 4; Length 121;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 860 GOKDAYG 866  
DB 97 GOKDAYG 103

## RESULT 49

US-09-270-767-33193  
; Sequence 33193, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 33193  
; LENGTH: 123  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-33193

Query Match 0.7%; Score 7; DB 4; Length 123;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 930 AAEAEOL 936  
DB 102 AAEAEOL 108

## RESULT 50

US-09-270-767-48410  
; Sequence 48410, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 48410  
; LENGTH: 123  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-48410

Query Match 0.7%; Score 7; DB 4; Length 123;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 930 AAEAEOL 936  
DB 102 AAEAEOL 108

## RESULT 51

US-09-270-767-57861  
; Sequence 57861, Application US/09270767  
; Patent No. 6703491

## ; GENERAL INFORMATION:

; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 57861  
; LENGTH: 136  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-57861

Query Match 0.7%; Score 7; DB 4; Length 136;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 701 STLQTT 707  
DB 10 STLQTT 16

## RESULT 52

US-09-134-000C-6044  
; Sequence 6044, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6044  
; LENGTH: 143  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-6044

Query Match 0.7%; Score 7; DB 4; Length 143;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 975 LPRTGSK 981  
DB 108 LPRTGSK 114

## RESULT 53

US-08-858-207A-523  
; Sequence 523, Application US/08858207A  
; Patent No. 6348328  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael  
; APPLICANT: Hodgson, John  
; APPLICANT: Knowles, David  
; APPLICANT: Nicholas, Richard  
; APPLICANT: Stodola, Robert  
; TITLE OF INVENTION: No. 6348328el Compounds  
; NUMBER OF SEQUENCES: 552  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Smithkline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 523:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-523

Query Match          0.7%; Score 7; DB 3; Length 144;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 503 EIKEGAF 509
   |||||
Db 45 EIKEGAF 51

RESULT 54
US-09-134-000C-6221
; Sequence 6221, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6221
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6221

Query Match          0.7%; Score 7; DB 4; Length 147;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 460 FONKTLR 466
   |||||
Db 5 FONKTLR 11

RESULT 55
US-08-679-493A-203
; Sequence 203, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
```

```
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: US/08/679,493A
; CURRENT FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 203
; LENGTH: 152
; TYPE: PRT
; ORGANISM: guineapig
US-08-679-493A-203

Query Match          0.7%; Score 7; DB 3; Length 152;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 803 TGLVEGK 809
   |||||
Db 35 TGLVEGK 41

RESULT 56
US-09-252-991A-24022
; Sequence 24022, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24022
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24022

Query Match          0.7%; Score 7; DB 4; Length 161;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 752 IAKAEKA 758
   |||||
Db 149 IAKAEKA 155

RESULT 57
US-09-538-092-450
; Sequence 450, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Ioic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
```



```
; SEQ ID NO 450
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YJL124C
US-09-538-092-450

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 172;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 952 SANSKDR 958
Db 2 SANSKDR 8

RESULT 58
US-09-270-767-38641
; Sequence 38641, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38641
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-38641

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 176;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 TKKATKN 767
Db 161 TKKATKN 167

RESULT 59
US-09-270-767-53858
; Sequence 53858, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53858
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-53858

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 176;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 TKKATKN 767
```

```
Db 161 TKKATKN 167

RESULT 60
US-09-893-737-22
; Sequence 22, Application US/09893737
; Patent No. 6822082
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-22

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 176;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 VKRLEKE 798
Db 114 VKRLEKE 120

RESULT 61
US-09-270-767-35060
; Sequence 35060, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35060
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-35060

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 180;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 575 AFLSNRL 581
Db 48 AFLSNRL 54

RESULT 62
US-09-270-767-50277
; Sequence 50277, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
```



; CURRENT APPLICATION NUMBER: US/09/305,984B  
; CURRENT FILING DATE: 1999-05-05  
; EARLIER APPLICATION NUMBER: 60/084,399  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: 09/305,984  
; EARLIER FILING DATE: 1999-05-05  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 215  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-305-984-18

Query Match 0.7%; Score 7; DB 3; Length 215;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522  
Db 199 TLELKD 205

RESULT 68  
US-09-305-984-64  
; Sequence 64, Application US/09305984B  
; Patent No. 6331407  
; GENERAL INFORMATION:  
; APPLICANT: No. 6331407ak, Rodger  
; APPLICANT: Toumanen, Elaine  
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME  
; FILE REFERENCE: 1340-1-016N1  
; CURRENT APPLICATION NUMBER: US/09/305,984B  
; CURRENT FILING DATE: 1999-05-05  
; EARLIER APPLICATION NUMBER: 60/084,399  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: 09/305,984  
; EARLIER FILING DATE: 1999-05-05  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 64  
; LENGTH: 215  
; TYPE: PRT  
; ORGANISM: Streptococcus Pneumoniae  
US-09-305-984-64

Query Match 0.7%; Score 7; DB 3; Length 215;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522  
Db 199 TLELKD 205

RESULT 69  
US-09-073-541A-18  
; Sequence 18, Application US/09073541A  
; Patent No. 6448224  
; GENERAL INFORMATION:  
; APPLICANT: No. 6448224ak, Rodger  
; APPLICANT: Toumanen, Elaine  
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME  
; FILE REFERENCE: 1340-1-016  
; CURRENT APPLICATION NUMBER: US/09/073,541A  
; CURRENT FILING DATE: 1998-05-06  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 215  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-073-541A-18

Query Match 0.7%; Score 7; DB 4; Length 215;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522  
Db 199 TLELKD 205

RESULT 70  
US-09-493-940-18  
; Sequence 18, Application US/09493940  
; Patent No. 6630583  
; GENERAL INFORMATION:  
; APPLICANT: No. 6630583ak, Rodger  
; APPLICANT: Toumanen, Elaine  
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME  
; FILE REFERENCE: 1340-1-016N1  
; CURRENT APPLICATION NUMBER: US/09/493,940  
; CURRENT FILING DATE: 2000-01-28  
; EARLIER APPLICATION NUMBER: 60/084,399  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: 09/305,984  
; EARLIER FILING DATE: 1999-05-05  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 215  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-493-940-18

Query Match 0.7%; Score 7; DB 4; Length 215;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522  
Db 199 TLELKD 205

RESULT 71  
US-09-493-940-64  
; Sequence 64, Application US/09493940  
; Patent No. 6630583  
; GENERAL INFORMATION:  
; APPLICANT: No. 6630583ak, Rodger  
; APPLICANT: Toumanen, Elaine  
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME  
; FILE REFERENCE: 1340-1-016N1  
; CURRENT APPLICATION NUMBER: US/09/493,940  
; CURRENT FILING DATE: 2000-01-28  
; EARLIER APPLICATION NUMBER: 60/084,399  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: 09/305,984  
; EARLIER FILING DATE: 1999-05-05  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 64  
; LENGTH: 215  
; TYPE: PRT  
; ORGANISM: Streptococcus Pneumoniae  
US-09-493-940-64

Query Match 0.7%; Score 7; DB 4; Length 215;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522  
Db 199 TLELKD 205

```
RESULT 72
US-09-583-110-5320
; Sequence 5320, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5320
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5320

Query Match          0.7%; Score 7; DB 4; Length 215;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKDK 522
|||||
Db 199 TLELKDK 205

RESULT 73
US-09-107-433-4810
; Sequence 4810, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; TITLE OF INVENTION: SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNO
; TITLE OF INVENTION: THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4810:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 219 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...219
; SEQUENCE DESCRIPTION: SEQ ID NO: 4810:
US-09-107-433-4810

Query Match          0.7%; Score 7; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKDK 522
|||||
Db 203 TLELKDK 209

RESULT 74
US-09-305-984-72
; Sequence 72, Application US/09305984B
; Patent No. 6331407
; GENERAL INFORMATION:
; APPLICANT: No. 6331407ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/305,984B
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 72
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-305-984-72

Query Match          0.7%; Score 7; DB 3; Length 224;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 516 TLELKDK 522
|||||
Db 208 TLELKDK 214

RESULT 75
US-09-305-984-74
; Sequence 74, Application US/09305984B
; Patent No. 6331407
; GENERAL INFORMATION:
; APPLICANT: No. 6331407ak, Rodger
; APPLICANT: Toumanen, Elaine
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
; FILE REFERENCE: 1340-1-016N1
; CURRENT APPLICATION NUMBER: US/09/305,984B
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,399
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: 09/305,984
; EARLIER FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 74
; LENGTH: 224
```

; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-305-984-74

Query Match 0.7%; Score 7; DB 3; Length 224;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522  
Db 208 TLELKD 214

RESULT 76  
US-09-305-984-76  
; Sequence 76, Application US/09305984B  
; Patent No. 6331407  
; GENERAL INFORMATION:  
; APPLICANT: Toumanen, Elaine  
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME  
; FILE REFERENCE: 1340-1-016N1  
; CURRENT APPLICATION NUMBER: US/09/305,984B  
; CURRENT FILING DATE: 1999-05-05  
; EARLIER APPLICATION NUMBER: 60/084,399  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: 09/305,984  
; EARLIER FILING DATE: 1999-05-05  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 76  
; LENGTH: 224  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae

US-09-305-984-76

Query Match 0.7%; Score 7; DB 3; Length 224;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522  
Db 208 TLELKD 214

RESULT 77  
US-09-493-940-72  
; Sequence 72, Application US/09493940  
; Patent No. 6630583  
; GENERAL INFORMATION:  
; APPLICANT: Toumanen, Elaine  
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME  
; FILE REFERENCE: 1340-1-016N1  
; CURRENT APPLICATION NUMBER: US/09/493,940  
; CURRENT FILING DATE: 2000-01-28  
; EARLIER APPLICATION NUMBER: 60/084,399  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: 09/305,984  
; EARLIER FILING DATE: 1999-05-05  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 72  
; LENGTH: 224  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae

Query Match 0.7%; Score 7; DB 4; Length 224;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522

Db 208 TLELKD 214

RESULT 78  
US-09-493-940-74  
; Sequence 74, Application US/09493940  
; Patent No. 6630583  
; GENERAL INFORMATION:  
; APPLICANT: No. 6630583ak, Rodger  
; APPLICANT: Toumanen, Elaine  
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME  
; FILE REFERENCE: 1340-1-016N1  
; CURRENT APPLICATION NUMBER: US/09/493,940  
; CURRENT FILING DATE: 2000-01-28  
; EARLIER APPLICATION NUMBER: 60/084,399  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: 09/305,984  
; EARLIER FILING DATE: 1999-05-05  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 74  
; LENGTH: 224  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae

US-09-493-940-74

Query Match 0.7%; Score 7; DB 4; Length 224;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522  
Db 208 TLELKD 214

RESULT 79  
US-09-493-940-76  
; Sequence 76, Application US/09493940  
; Patent No. 6630583  
; GENERAL INFORMATION:  
; APPLICANT: No. 6630583ak, Rodger  
; APPLICANT: Toumanen, Elaine  
; TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME  
; FILE REFERENCE: 1340-1-016N1  
; CURRENT APPLICATION NUMBER: US/09/493,940  
; CURRENT FILING DATE: 2000-01-28  
; EARLIER APPLICATION NUMBER: 60/084,399  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: 09/305,984  
; EARLIER FILING DATE: 1999-05-05  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 76  
; LENGTH: 224  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae

US-09-493-940-76

Query Match 0.7%; Score 7; DB 4; Length 224;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 TLELKD 522  
Db 208 TLELKD 214

RESULT 80  
US-09-902-540-11067  
; Sequence 11067, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:

```
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 11067
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-11067

Query Match          0.7%; Score 7; DB 4; Length 236;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 143 GNTLVGL 149
Db 96 GNTLVGL 102

RESULT 81
US-09-198-452A-399
; Sequence 399, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 399
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-399

Query Match          0.7%; Score 7; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 595 IPVQAPS 601
Db 112 IPVQAPS 118

RESULT 82
US-09-438-185A-384
; Sequence 384, Application US/09438185A
; Patent No. 6822071
; GENERAL INFORMATION:
; APPLICANT: Stephens, Richard
; APPLICANT: Mitchell, Wayne
; APPLICANT: Kalman, Sue
; APPLICANT: Davis, Ronald
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
; FILE REFERENCE: 018941-000411US
; CURRENT APPLICATION NUMBER: US/09/438,185A
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: US 60/108,279
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: US 60/128,606
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 1074
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 384
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; OTHER INFORMATION: CPn0382
US-09-438-185A-384

Query Match          0.7%; Score 7; DB 4; Length 239;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 595 IPVQAPS 601
Db 112 IPVQAPS 118

RESULT 83
US-08-867-087B-11
; Sequence 11, Application US/08867087B
; Patent No. 5990386
; GENERAL INFORMATION:
; APPLICANT: An, Gynheung
; TITLE OF INVENTION: GENES CONTROLLING FLORAL DEVELOPMENT
; TITLE OF INVENTION: AND APICAL DOMINANCE IN PLANTS
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist Sparkman Campbell Leigh &
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,087B
; FILING DATE: June 2, 1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/323,449
; FILING DATE: October 14, 1994
; APPLICATION NUMBER: U.S. 08/485,981
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Alan. E.
; REGISTRATION NUMBER: 35,123
; REFERENCE/DOCKET NUMBER: 4630-47071
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acid residues
; TYPE: amino acid
; TOPOLOGY: linear
US-08-867-087B-11

Query Match          0.7%; Score 7; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 704 RQTTQTQ 710
Db 194 RQTTQTQ 200
```

RESULT 84  
US-09-107-532A-5876  
; Sequence 5876, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; Filing DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; Filing DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; Filing DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 5876:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 241 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (B) LOCATION 1...241  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5876:  
US-09-107-532A-5876  
Query Match 0.7%; Score 7; DB 4; Length 241;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 793 KRLEKEL 799  
Db 126 KRLEKEL 132  
RESULT 85  
US-09-252-991A-21887  
; Sequence 21887, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; Filing DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR Filing DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR Filing DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32999  
; LENGTH: 248  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32999  
Query Match 0.7%; Score 7; DB 4; Length 248;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 717 AGKALLS 723

; PRIOR Filing DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR Filing DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21887  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21887  
Query Match 0.7%; Score 7; DB 4; Length 242;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 882 LAVATLA 888  
Db 27 LAVATLA 33  
RESULT 86  
US-09-328-352-4519  
; Sequence 4519, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT Filing DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4519  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4519  
Query Match 0.7%; Score 7; DB 4; Length 243;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 254 LPESLET 260  
Db 63 LPESLET 69  
RESULT 87  
US-09-252-991A-32999  
; Sequence 32999, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; Filing DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR Filing DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR Filing DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32999  
; LENGTH: 248  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32999  
Query Match 0.7%; Score 7; DB 4; Length 248;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 717 AGKALLS 723



```
Db      118 AGKALLS 124
|||||
RESULT 88
US-09-270-767-40859
; Sequence 40859, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40859
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-40859

Query Match      0.7%; Score 7; DB 4; Length 250;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      256 ESLETIS 262
|||||
Db      65 ESLETIS 71

RESULT 89
US-09-270-767-56075
; Sequence 56075, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56075
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-56075

Query Match      0.7%; Score 7; DB 4; Length 250;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      256 ESLETIS 262
|||||
Db      65 ESLETIS 71

RESULT 90
US-09-248-796A-17359
; Sequence 17359, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17359
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17359

Query Match      0.7%; Score 7; DB 4; Length 256;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      684 TIVDEK 690
|||||
Db      67 TIVDEK 73

RESULT 91
US-09-489-039A-7271
; Sequence 7271, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7271
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7271

Query Match      0.7%; Score 7; DB 4; Length 259;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      750 KAIKAE 756
|||||
Db      59 KAIKAE 65

RESULT 92
US-09-489-039A-11968
; Sequence 11968, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11968
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11968

Query Match      0.7%; Score 7; DB 4; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      378 NPSGLAT 384
|||||
Db      215 NPSGLAT 221
```

```
RESULT 93
US-09-270-767-42555
; Sequence 42555, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42555
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-42555

Query Match      0.7%; Score 7; DB 4; Length 268;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 701 STLQTT 707
   |||||
Db 142 STLQTT 148

RESULT 94
US-09-248-796A-19421
; Sequence 19421, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19421
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19421

Query Match      0.7%; Score 7; DB 4; Length 270;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 900 NSKLSQL 906
   |||||
Db 137 NSKLSQL 143

RESULT 95
US-09-489-039A-10410
; Sequence 10410, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
```

```
; SEQ ID NO 10410
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10410

Query Match      0.7%; Score 7; DB 4; Length 272;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 992 TSVALLS 998
   |||||
Db 18 TSVALLS 24

RESULT 96
US-08-741-437-1
; Sequence 1, Application US/08741437
; Patent No. 5843665
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN PYROPHOSPHATASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/741,437
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0148 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 855-0555
; TELEFAX: (415) 845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
US-08-741-437-1

Query Match      0.7%; Score 7; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 757 KALVTKK 763
   |||||
Db 228 KALVTKK 234

RESULT 97
US-09-134-593-1
; Sequence 1, Application US/09134593
```

Patent No. 5981232  
; GENERAL INFORMATION:  
; APPLICANT: Hawkins, Phillip R.  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: A NOVEL HUMAN PYROPHOSPHATASE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/134,593  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/741,437  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0148 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 855-0555  
; TELEFAX: (415) 845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 289 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY:  
; CLONE: Consensus  
US-09-134-593-1

Query Match 0.7%; Score 7; DB 2; Length 289;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 757 KALVTKK 763  
Db 228 KALVTKK 234

RESULT 98  
US-09-248-796A-18420  
; Sequence 18420, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 18420  
; LENGTH: 289  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-18420

Query Match 0.7%; Score 7; DB 4; Length 289;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 763 KATKNGQ 769  
Db 195 KATKNGQ 201

RESULT 99  
US-09-538-092-1335  
; Sequence 1335, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuratSeqFormatter Version 0.9  
; SEQ ID NO 1335  
; LENGTH: 289  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number Q15181  
US-09-538-092-1335

Query Match 0.7%; Score 7; DB 4; Length 289;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 757 KALVTKK 763  
Db 228 KALVTKK 234

RESULT 100  
US-09-535-521-2  
; Sequence 2, Application US/09535521  
; Patent No. 6410714  
; GENERAL INFORMATION:  
; APPLICANT: Weber, Eric R.  
; APPLICANT: McCall, Catherine A.  
; TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IGE RECEPTOR (CANINE CD23)  
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF  
; FILE REFERENCE: AL-5  
; CURRENT APPLICATION NUMBER: US/09/535,521  
; CURRENT FILING DATE: 2000-03-24  
; EARLIER APPLICATION NUMBER: 60/125,913  
; EARLIER FILING DATE: 1999-03-24  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 292  
; TYPE: PRT  
; ORGANISM: Canis familiaris  
US-09-535-521-2

Query Match 0.7%; Score 7; DB 4; Length 292;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 627 LKQLEVA 633

Db           |||||||  
          57 LKQLEVA 63

Search completed: August 28, 2005, 11:17:45  
Job time : 46 secs

**THIS PAGE BLANK (USPTO)**